华章数学译丛

38

Real Analysis and Probability

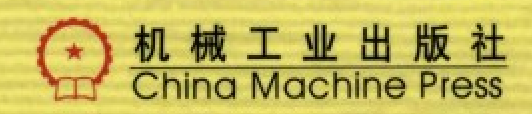
(Second Edition)

实分析和概率论

（原书第2版）

（美）R. M. Dudley 麻省理工学院 著

赵选民 孙浩 译



（原书第2版）

实分析和概率论

本书在两个方面获得了极佳的成功。一是它是一本全面、新颖的实分析教程，二是它是一本数学理论完整和自成体系的概率论教程。本书无疑给出了一种严谨和完整的新标准。

——美国数学会公报

这是一本非凡的著作。在教学和参考两个方面，本书将成为一本标准化教材，它全面地介绍了实分析的必备知识，且证明贯穿全书。书中的一些主题和证明极少在其他教科书中见到。

——爱丁堡数学会学报

严谨、精深、新颖，这是一本适用于数学专业研究生的教材。

——ISI的简短书评

这是一本广受称赞的教科书，清晰地讲解了现代概率论以及概率测度与度量空间之间的相互关系。本书分为两部分：第一部分介绍了实分析的内容，包括基础集合论、一般拓扑、测度、积分、巴拿赫空间及希尔伯特空间上的泛函分析、凸集和函数，以及拓扑空间上的测度。第二部分介绍了基于测度论的概率方面的内容，包括大数定律、遍历定理、中心极限定理、条件期望、鞅收敛。另外，随机过程一章介绍了布朗运动以及布朗桥。

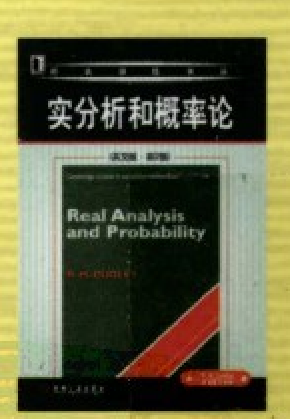
与前版相比，本版内容更加丰富、完整，包括了实数系的基础知识和代数中一致逼近的斯通－魏尔斯特拉斯定理，修订和改进了几节的内容，扩充了大量历史注释，添加了一些新的习题，以及习题的解题提示。

作者简介

R. M. Dudley 麻省理工学院数学系教授。除本书外，他还著有《Differentiability of Six Operators on Nonsmooth Functions and p-Variation》、《Uniform Central Limit Theorems》等书。

Real Analysis and Probability

(Second Edition)



影印版
ISBN 7-111-19348-2
定价：69.00元

投稿热线：(010) 88379604
购书热线：(010) 68995259, 68995264
读者信箱：hzjsj@hzbook.com

华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

封面设计·杨宇梅



上架指导：数学


ISBN 978-7-111-23480-7
9 787111 234807

定价：55.00 元

华章数学译丛

38

Real Analysis and Probability
(Second Edition)

实分析和概率论

(原书第2版)

(美) R. M. Dudley 著
麻省理工学院
赵选民 孙浩 译

机械工业出版社
China Machine Press

本书清晰地讲解了现代概率论以及概率测度与度量空间之间的相互关系．本书分两部分，第一部分介绍了实分析的内容，包括基础集合论、一般拓扑、测度、积分、巴拿赫空间及希尔伯特空间上的函数分析、凸集和函数以及拓扑空间上的测度，第二部分介绍了基于测度论上的概率论，包括大数定律、遍历定理、中心极限定理、条件期望、鞅收敛．另外，随机过程一章介绍了布朗运动以及布朗桥．

本书适合于概率论与数理统计方向的研究生，以及与之相关的研究生阅读，也适合于数学系高年级学生以及数学研究工作者参考使用．

R. M. Dudley：Real Analysis and Probability，Second Edition（ISBN 0-521-80972-X）.

Originally published by Cambridge University Press in 2003.

This Chinese edition is published with the permission of the Syndicate of the Press of the University of Cambridge，Cambridge，England.

本书版权登记号：图字：01-2006-2852

图书在版编目（CIP）数据

实分析与概率论（原书第2版）/（美）达德利（Dudley，R. M.）著；赵选民，孙浩译．—北京：机械工业出版社，2008.6

（华章数学译丛）

书名原文：Real Analysis and Probability，Second Edition

ISBN 978-7-111-23480-7

Ⅰ.实…　Ⅱ.①达…　②赵…　③孙…　Ⅲ.①实分析　②概率论　Ⅳ.O174.1　O211

中国版本图书馆CIP数据核字（2008）第054804号

机械工业出版社（北京市西城区百万庄大街22号　邮政编码　100037）

责任编辑：王春华

山西新华印业有限公司新华印刷分公司印刷·新华书店北京发行所发行

2008年6月第1版第1次印刷

186mm×240mm·24印张

标准书号：ISBN 978-111-23480-7

定价：55.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：（010）68326294

译 者 序

2007年2月4日，我尊敬的导师赵选民教授(译者之一)带着未能完成该书翻译的遗憾，永远地离开了爱着他的学生和他心爱的工作岗位．本书现在顺利出版，特以告慰赵老师的在天之灵．

本书主要包含两部分内容：实分析与概率论．这是一本非常系统的、完整的以实分析中的测度论理论为基础，揭示概率论的规律与理论的宏著．该论著主要适合于概率论与数理统计方向的研究生，以及与之相关方向的研究生阅读，也适合于数学系高年级学生以及数学研究工作者参考使用．

本书是在机械工业出版社华章分社大力支持下翻译完成的，翻译工作历时一年两个月．在翻译中，主要强调在尊重原文的基础上，数学专有名词翻译的规范性以及数学论证的逻辑性．译著经过多次讨论与修改，才形成现有状态．但由于译者自身知识水平有限，瑕疵在所难免，希望各位同仁不吝赐教，以便修正．

本书的翻译工作主要由赵选民教授和孙浩博士负责，刘向增、张燕、李颖、刘威、王彩玲、丛方媛、牛玉俊和何华等同志参与了讨论、校对、整理和编辑等相关工作，机械工业出版社华章分社的编辑们对译文提出了许多中肯的修订意见．在此向所有支持、参与本书出版的人士表示衷心的感谢，也将此译著献给 R. M. Dudley 教授，感谢他为我们系统学习测度论基础上的概率论提供了范本．

孙　浩

2008年4月25日于西安



前　言

本书是为研究生一年级学生而编写的．对欧几里得空间中的媒介分析的了解将对学习本书有很大的帮助，但是在本书中我并未将它列为预备知识．为了使这本书更加完整，我在第 1 章里增加了实数系和斯通-魏尔斯特拉斯定理，删除了一些其他书中已有的证明，并保留了像附录 B 中的复变理论方面的资料．

第 1 ~ 5 章提供的实分析可供一学期学习．之后的内容是可供一学期学习的概率论，这包括第 8 ~ 10 章，以及第 11 章、第 12 章的部分内容．标有 * 的段落和章节不常用到，至少是在以后很少用到，所以在第一次阅读时可以跳过，我们可以在以后用到时再读，这包括第 6 章中的一些和第 7 章的大部分内容．

一些不太重要的证明留给读者作为练习．如果读者发现本书的一些漏洞或者错误还请告知，我会很高兴地加以改进修正．虽然我认真地检查了所有的习题和提示，但是错误在所难免．提示也可能会起到误导的作用，所以请读者先尝试着自己做习题，尽可能不要看提示．

我尽自己最大的努力给出定理的最直接、最简明的证明．关于度量空间的完备化、强大数定律、遍历定理、鞅收敛定理、次加性遍历定理、Hartman-Wintner 重对数律的证明在相应的期刊中可以找到．

1950 年前后，当 Halmos 经典测度论问世的时候，测度论更高级的观点——局部紧空间也出现了，即本书 7.3 节中的内容．之后，概率论研究的主要方向转移到度量空间．第 11 章给出了较前沿的度量和概率之间联系的一些事实．附录 E 指出了在非度量(局部)紧空间中测度可能出错的地方．这部分内容仅仅经过一年的学习是不能很好地理解的，这需要一个长期的过程，但本书为眼前和以后的研究介绍了一些本质的内容．

在每章末都有习题，习题按照由简到难的顺序给出，并且多数习题都有解题提示，本版书中增加或改进了一些提示．

我也尽量给出书中定理的出处，有时这些历史注释和参考文献会很多，我都把它们罗列在各章的后面．有些注释被引申，有的被更正．但是，书中有关这些定理历史发展的评论仅是一家之言．

本书从 1967 年在麻省理工学院开始到 1976 年在丹麦奥尔胡斯完成．非常感谢 Ken. Alexander、Deborah Allinger、Laura Clemens、Ken Davidson、Don Davis、Persi Diaconis、Arnout Eikeboom、Sy Friedman、David Gillman、José Gonzalez、E. Griffor、Leonid Grinblat、Dominique Haughton、J. Hoffmann-Jørgensen、Arthur Mattuck、Jim Munkres、R. Proctor、Nick Reingold、Rae Shortt、Dorothy Maharam Stone、Evangelos Tabakis、Jin-Gen Yang，以及其他的同事和学生，感谢他们提出的宝贵建议．

在第 1 版出版时，Ken Brown 给出很多有益的建议，在此表示感谢，还要感谢 Justin Corvino、Charles Goldie、Charles Hadlock、Michael Jansson、Suman Majumdar、Rimas Norvaiša、Mark Pinsky、Andrew Rosalsky. Rae Shortt 和 Dewey Tucker. 特别需要感谢的是 Andries Lenstra、Valentin Petrov，他们给出了许多更好的建议．我们主要对 10.2 节的正则条件概率及第 12 章的马尔可夫时作了修改．

R. M. Dudley

目　录

第 1 章　基础知识：集合论

在建造房子时，建设者会用与房子其他部分不一样的材料和方法来打造地基. 同样，几乎每一个数学分支都以公理集合论作为其基础. 这个基础被大多数关注基础理论的逻辑学家和数学家所接受，但是只有少数数学家有时间或意愿去详细研究公理集合论.

做另一个比喻，高级计算机语言和用它们编写的程序建立了计算机硬件和软件的基础. 但是，编写高级计算机程序的人需要了解多少计算机硬件和操作系统的知识则取决于他手边的问题.

在现代实分析中，集合论的问题比以前代数、复分析、几何和应用数学中的问题要多. 例如，在实分析中相对较近发展的"非标准分析"允许正数可以无限小但不为零. 非标准分析比早期发展的实分析更强地依赖于集合论的特性.

本章将介绍本书后面要用到的一些集合论符号和概念. 换句话说，本章只给出最基本的集合论知识. 附录 A 较详细地介绍了集合论，包括一些集合论公理，但是本书不打算介绍非标准分析或者更深入地讨论集合论.

本章所定义的许多概念在以后的章节里都要用到，希望读者能熟记.

1.1　集合论的定义和实数系

定义至少有两个目的. 首先，就像一本普通字典一样，定义试图给出见解，传达一种思想，或者用熟悉的概念去解释陌生的概念，但并不详尽说明或彻底研究所定义单词的全部意义. 我们称这种定义为*非形式*(informal)定义. 在大多数数学和其他科学领域中，*形式*(formal)定义是完全准确的，因此，人们可以科学地判断一个有关命题的真伪. 在形式定义中，一个熟悉的术语(例如，普通的长度单位或数字)可以用不熟悉的术语定义. 集合论中的大多数定义都是形式的. 其次，集合论的另一个目的是不但为自己也为所有数学分支提供清晰的逻辑结构. 由此就产生从哪里开始定义的问题.

非形式的字典定义常常由一些同义词组成. 例如，字典中用"high"和"tall"相互定义，这种定义方法对于知道其中一个词的人来说是有帮助的，但是对于一个通过字典学习英语的人来说，这样的定义是无用的. 这种情况在一定程度上反映了人们在学习中所遇到的困难，因为字典中所有单词都是用其他单词来定义的. 因此，人们在开始时应至少弄清字典中一些单词的意思，而不是在用到时才去翻字典.

有些单词(例如，"and"、"or"和"but"这三个连词)非常常见，但很难用另外的词来定义. 这时，我们可能通过一些规则来定义含有连词的句子的含义，其中用这些连词连接起来的单词或短语的词义已经给定.

乍一想，你也许认为集合论中最重要的定义是"集合"，但是恰恰相反，因为数学的整个逻辑结构归约为集合或者由集合来定义，因此不能给出它的一个形式的、准确的定义. 相反，有一些规则(公理、推理规则等)能起到定义集合的作用. 一个初步的、非形式的集合(set)定义是"任何数学对象的全体"，但是随着学习的深入我们在对这个定义不断地加以明确和修正.

在某些方面，定义集合的问题类似于定义数(number)的问题. 经过几年的学习之后，学生们知

道了0，1，2，…这些数，并且知道了它们之间的运算规律．但是许多人还是很难确切地说明什么是数．即使人们完全同意算术规律，但是不同的人可能会给出数字1不同的定义．

在19世纪后期，数学家开始研究数的精确定义．一种方法是从0开始，通过取“后继者”或“下一个更大的整数”的方法来产生更多的数．

2

如果定义了0，那么0之后的数字也就确定了．如果用n'表示n之后的数字，那么我们用0及它的后继者来定义一个序列0，$0'$，$0''$，$0'''$，…我们可以得到通常的整数定义．为了做到这一点，我们将在等号之前加上冒号“:=”表示“定义为相等”，例如$1:=0'$，$2:=0''$，$3:=0'''$，$4:=0''''$，等等．这样一直做下去，这些定义是精确的．这样，人们可以得到一本厚厚的数的字典，相当精确（尽管不太实用）但还不完整，因为0和后继运算没有被形式定义．大多数数系结构可以通过给定0和后继者的规则来确定．例如，一种规则是：假若$m'=n'$，则$m=n$.

一旦有足够多的规则来定义非负整数的结构，那么重要的是结构而不是结构中的每个具体的元素．

简言之，如果我们想要尽可能精确地建立严格的数学逻辑结构，那么非形式的定义不能是该结构的一部分，尽管在有些情况下，非正式的定义会帮助人们解释结构．然而，我们至少还要留下一些未被定义的基本概念．本节我们将给出公理体系和其他规则，并用基本概念定义其余的概念．

下面我们再次给出集合的非形式定义：一个集合是任何对象的全体．在数学上，这种对象将是数学对象，例如数、点、向量，或者其他的集合．（实际上，从集合论的观点来看，所有的数学对象都是一种集合或者另一种集合．）如果对象x是集合y中的元素，记作“$x\in y$”，读作“x属于y”或者“x在y中”．如果集合S是一个有限集，即S中的元素可以排成一个有限列x_1，x_2，…，x_n，那么S可以写作$S=\{x_1, x_2, \cdots, x_n\}$．例如，$\{2, 3\}$是只有2和3两个元素的集合．其中记号“$\in$”也是没有被正式定义的几个基本记号之一．

一个集合可以只有一个元素，例如，只有一个元素x的集合，记作$\{x\}$，读作“单元素集x”．在集合论中，$\{x\}$与x本身是不同的．例如，如果$x=\{1, 2\}$，则x有两个元素，而$\{x\}$只有一个元素．

集合A包含于(included)集合B中，或者A是B的子集(subset)，记作$A\subset B$，当且仅当A中的每一个元素都是集合B中的元素．一个等价的说法是，集合B包含(include)集合A，记作$B\supset A$．集合B包含(contain)x是指$x\in B$．当$B\supset A$，许多人也称B包含了A．

“if and only if”有时简记为“iff”．例如，$A\subset B$ iff 对所有x，如果$x\in A$，则$x\in B$．

集合论中一个重要的规则称为“外延性”．它是说，如果两个集合A和B有完全相同的元素，即对任意的元素x，$x\in A$当且仅当$x\in B$，或者等价地$A\subset B$且$B\subset A$，那么两个集合相等，记作$A=B$．例如，$\{2, 3\}=\{3, 2\}$．集合中元素之间的次序可以是不同的，而它们的元素是相同的．在某种意义上，外延性是集合相等的一个定义．在集合论中另一种比较常用的观点是，任意两个对象是相等的当且仅当它们是完全同一的．因此，“$\{2, 3\}$”和“$\{3, 2\}$”是名字不同的同一个集合．

3

外延性也可以用来给出集合的非形式定义．一个集合可以仅以其所含的元素来定义，但元素之间的结构和关系与集合的定义无关．

与把有限个元素一一罗列出来不同，定义一个特殊集合的主要方法是给出元素所满足的条件．记号$\{x: \cdots\}$表示集合中的所有x满足…例如，集合$\{x: (x-4)^2=4\}=\{2, 6\}=\{6, 2\}$．

一种惯用的方法是在某一符号上划一道线表示“非”，例如，$a\neq b$表示“a不等于b”，符号“$\notin$”表示“不是…中的元素”．因此，$x\notin y$表示x不是y中的元素，例如$3\notin\{1, 2\}$．

由条件定义集合如果不仔细可能导致矛盾．例如，令 $r=\{x: x\notin x\}$，则 $r\notin r$ 意味着 $r\in r$，反之也成立(这就是罗素(Bertrand Russell)悖论)．这个悖论可以通过限制某些集合的条件而得以避免．因此，$\{x\in A: \cdots x\cdots\}$ 表示"A 中满足条件 $\cdots x\cdots$ 的所有 x 的集合"．当 A 是一个已知的集合时，只要使用这种形式的定义，用此方法就能定义新的集合，并且不会出现矛盾．

不管怎样，一个集合是它自己的元素看起来似乎有点奇怪，我们将在附录 A 中给予证明(定理 A.1.9)．从附录 A 给出的集合论的公理知，没有集合是它自己的元素．从这个意义上讲，在罗素悖论中命名的集合 r 就是所有集合的全体，在集合论中有时称之为"全域"．这里集合的非形式定义(即集合是任何对象的全体)确实是不严格的．附录 A 中的公理给出了一些条件，在该条件下某些全体可能是集合也可能不是集合．例如，全域不是一个集合．

数学上人们常常关心一个固定点集 y 的内部情况．我们用 $\{x: \cdots x\cdots\}$ 来表示 $\{x\in y: \cdots x\cdots\}$．

下面给出集合论中各种运算的定义．在这里我们不明确指明集合中的元素，公理就意味着它们是集合中的元素(见附录 A)．

没有任何元素的集合称为"空集"，记作 $\varnothing$，即对所有的 x，都有 $x\notin\varnothing$．由外延性，这个集合是唯一的．如果 B 是任意一个集合，则 2^B 称为 B 的"幂集"，是 B 的所有子集构成的集合．例如，如果集合 B 含有 3 个元素，则 2^B 有 $2^3=8$ 个元素．同时，$2^{\varnothing}=\{\varnothing\}\neq\varnothing$．

4

$A\cap B$ 称为 A 和 B 的交集，定义为 $A\cap B:=\{x\in A: x\in B\}$．换句话说，$A\cap B$ 是所有既属于 A 又属于 B 的 x 所组成的集合．$A\cup B$ 称为 A 和 B 的并集，表示对任意的 x，$x\in A\cup B$ 当且仅当 $x\in A$ 或 $x\in B$(或同时属于两者)．$A\setminus B$(读作"A 减 B")，表示所有在 A 中而不在 B 中的 x 所组成的集合，有时称为(B 在 A 中的)相对补集(relative complement)．对称差(symmetric difference) $A\Delta B$ 定义为 $(A\setminus B)\cup(B\setminus A)$．

用 $\mathbb{N}$ 表示所有非负整数 0，1，2，…的集合．(形式上，非负整数通常通过定义 0 为空集 $\varnothing$，1 为 $\{\varnothing\}$ 来定义，一般地，后继的数由上面提到的运算定义 $n'=n\cup\{n\}$ 而得到，更详细的说明参见附录 A.)

有序对(ordered pair)非形式地定义为由一对给定次序的数学对象组成，例如 $\langle x, y\rangle$，这里 x 称为有序对 $\langle x, y\rangle$ 的第一个元素，y 为第二个元素．有序对满足下面的公理：对所有的 x，y，u 和 v，$\langle x, y\rangle=\langle u, v\rangle$ 当且仅当 $x=u$，$y=v$. 在有序对 $\langle x, y\rangle$ 中可以出现 $x=y$ 的情况．有序对可以形式地通过满足公理的(无序的、平凡的)集合来定义．通常的方法是令 $\langle x, y\rangle:=\{\{x\}, \{x, y\}\}$(见附录 A)．由外延性知，$\{\{x\}, \{x, y\}\}=\{\{y, x\}, \{x\}\}$．

数学中的主要概念之一是函数．函数的非形式定义是，给定两个集合 D 和 E，D 上的函数定义为：对于 D 中的每一个元素 x，在 E 中有唯一的元素 $f(x)$ 和它对应．函数的形式定义是，一个函数(function)定义为有序对 $\langle x, y\rangle$ 的一个集合 f，使得对于任意的 x，y 和 z，如果 $\langle x, y\rangle\in f$ 且 $\langle x, z\rangle\in f$，则 $y=z$. 例如，$\{\langle 2, 4\rangle, \langle -2, 4\rangle\}$ 是一个函数，但是 $\{\langle 4, 2\rangle, \langle 4, -2\rangle\}$ 不是函数．函数的有序对的集合非形式地称为函数的图形(graph)(对于 $D=E=\mathbb{R}$ 的情况，$\mathbb{R}$ 是实数集)．

函数 f 的定义域(domain)是使得对某些 y，$\langle x, y\rangle\in f$ 的所有 x 组成的集合，记作 $\mathrm{dom}\, f$. 由函数的定义，y 是唯一确定的，记为 $f(x)$．f 的值域(range)是使得对于某些 x，$f(x)=y$ 的所有 y 的集合，记作 $\mathrm{ran}\, f$.

具有定义域 A 并且值域包含在集合 B 中的函数 f 称为是定义在 A 上的或是从 A 映射到 B 的．如果 f 的值域等于 B，则称 f 是映上(onto)到 B 的．

符号“$\mapsto$”有时用来表示或定义函数，一个函数可以写为“$x \mapsto f(x)$”. 例如，“$x \mapsto x^3$”或者“f: $x \mapsto x^3$”是指对(f的定义域中的)所有 x，使得 $f(x)=x^3$ 的函数 f. 为明确定义域，通常使用与之相关的记号“f: $A \mapsto B$”，再加上 f 精确的定义，就表示它定义为从 A 映射到 B 的函数(而不是指 $f(A)=B$，为了区分记号 $\mapsto$ 的这两个应用，把 A 和 B 写成大写字母，把它们的元素写为小写字母，比如 x).

5

如果 X 是任意一个集合，A 是 X 的任意一个子集，A(在 X 上)的示性函数(indicator function)定义为

$$1_A(x) := \begin{cases} 1, & 若\ x \in A \\ 0, & 若\ x \notin A. \end{cases}$$

(许多数学家称这个函数为 A 的特征函数(characteristic function.)在概率论中，“特征函数”是指傅里叶变换，这将在第 9 章介绍.)

一个序列(sequence)是定义在$\mathbb{N}$或者所有正整数集合$\{1, 2, 3, 4, \cdots\}$上的函数. 对所有的 n，满足 $f(n)=x_n$ 的序列 f 通常写为 $\{x_n\}_{n\geq 1}$.

在形式场合下，每个集合都是集合的集合(集合的每个元素也是一个集合). 如果一个集合被非形式地看作是由一些集合所组成的，那么我们称之为集族、集类，或者集合的全体. 设$\mathcal{V}$是一个集族，那么$\mathcal{V}$的并集(union)定义为

$$\bigcup \mathcal{V} := \{x: 对某些\ A \in \mathcal{V}, x \in A\}.$$

同样，非空集族$\mathcal{V}$的交集(intersection)定义为

$$\bigcap \mathcal{V} := \{x: 对所有的\ A \in \mathcal{V}, x \in A\}.$$

于是对任意两个集合 A 和 B，$\cup\{A, B\}=A\cup B$，$\cap\{A, B\}=A\cap B$. 诸如$\cup\mathcal{V}$和$\cap\mathcal{V}$的记号在集合论中是最常用的. 在数学的其他分支中，多于两个的集合的交集和并集经常用下标表示. 如果$\{A_n\}_{n\geq 1}$是一个集合序列，则它们的并集写为

$$\bigcup_n A_n := \bigcup_{n=1}^{\infty} A_n := \{x: 对某些\ n, x \in A_n\}.$$

同样地，它们的交集写为

$$\bigcap_{n\geq 1} A_n := \bigcap_{n=1}^{\infty} A_n := \{x: 对所有的\ n, x \in A_n\}.$$

有限个集合 A_1，$\cdots$，A_n 的并集写为

$$\bigcup_{1\leq i\leq n} A_i := \bigcup_{i=1}^{n} A_i := \{x: 对某些\ i = 1, \cdots, n, x \in A_i\}.$$

6

类似地，将并改为交，“一些”改为“所有的”，即可定义有限个集合的交集.

更一般地，令 I 是任意一个集合，设 A 是定义在 I 上的函数，其值是集合 $A_i := A(i)$，那么所有这些集合 A_i 的并集写为

$$\bigcup_i A_i := \bigcup_{i\in I} A_i := \{x: 对某些\ i, x \in A_i\}.$$

在这种情况下，集合 I 称为指标集(index set)，这意味着 I 是函数 $i \mapsto A_i$ 的定义域. 如果从上下文可以清楚地看出 I 是什么，就像上面的第一个表示式那样，指标集在记号中可以被省略. 同样地，交集可以写为

$$\bigcap_i A_i := \bigcap_{i \in I} A_i := \{x; \text{对所有的 } i \in I, x \in A_i\},$$

这里的 I 通常是一个非空集合．但当这些集合是所讨论的某一固定集合（比如说 X）的一切子集时，就出现例外．假定 $t \notin I$ 并且令 $A_t := X$，如果 I 是非空集合，那么用 $I \cup \{t\}$ 代替 I 不会改变 $\bigcap_{i \in I} A_i$．在 I 是空集的情况下，我们可以令 $\bigcap_{i \in \varnothing} A_i = X$.

下面介绍在数理逻辑中两个用得比较多的缩写符号：$\forall$ 表示"对任意的"，而 $\exists$ 表示"存在"．例如，$(\forall x \in A)(\exists y \in B)\cdots$ 表示对任意的 $x \in A$，在 B 中存在一个 y，使得$\cdots$.

两个集合 A，B 称为不相交的（disjoint），当且仅当 $A \cap B = \varnothing$．集合 A_i 对 $i \in I$ 称为是不相交的，当且仅当对 I 中所有的 $i \neq j$，$A_i \cap A_j = \varnothing$.

下面将给出不同数域的定义，首先是实数域．假定所有的读者都熟悉整数和有理数．更详细的介绍参见附录 A.4.

回忆以前学过的：$\mathbb{N}$ 表示所有非负整数 0，1，2，$\cdots$ 组成的集合，$\mathbb{Z}$ 表示所有整数 0，± 1，± 2，$\cdots$ 组成的集合，$\mathbb{Q}$ 表示所有有理数 m/n 组成的集合，其中 $m \in \mathbb{Z}$，$n \in \mathbb{Z}$，并且 $n \neq 0$.

实数可以用不同的方法来定义．一种常用的方法是用十进制数表示：x 是实数当且仅当 $x = \pm y$，其中 $y = n + \sum_{j=1}^{\infty} d_j/10^j (n \in \mathbb{N})$，并且每个数字 d_j 是 0～9 之间的整数．但是这种表示在分析证明中用起来不方便，而且对形如 $m/10^k (m \in \mathbb{Z}, m \neq 0, k \in \mathbb{N})$ 的有理数，表示是不唯一的．还可以根据一般的收敛到实数 x 的有理数列来定义实数 x，当讲完度量空间后，我们将在 2.5 节研究这一问题.

这里所用的实数的形式定义是用戴德金（Dedekind）分割法来定义，定义如下：一个分割（cut）是一个集合 $C \subset \mathbb{Q}$，使得 $C \notin \varnothing$；$C \neq \mathbb{Q}$；无论何时 $q \in C$，如果 $r \in \mathbb{Q}$，且 $r < q$，则 $r \in C$，并且存在 $s \in \mathbb{Q}$，满足 $s > q$ 且 $s \in C$.

7

令 $\mathbb{R}$ 表示所有实数的集合．那么形式上 $\mathbb{R}$ 就是所有分割的集合．也可非形式地将实数 x 与分割 C 之间的一一对应记作 $C_x = \{q \in \mathbb{Q} : q < x\}$，其中 $C = C_x$ 或者 $x = x_C$.

实数中的次序 $x \leqslant y$，根据分割定义为 $C_x \subset C_y$．一个实数集 E 称为上有界（bounded above），它的上界（upper bound）为 y，当且仅当对任意的 $x \in E$，有 $x \leqslant y$．称 y 为 E 的上确界（supremum）或者最小上界（least upper bound），记作 $y = \sup E$，当且仅当 y 是一个上界并且对 E 的任意一个上界 z，有 $y \leqslant z$．关于 $\mathbb{R}$ 的一个基本事实是对每一个非空集 $E \subset \mathbb{R}$，使得 E 是上有界的，则上确界 $y = \sup E$ 存在．这用分割法很容易证明：对所有的 $x \in E$，令 C_y 是 C_x 的并集，证明见附录 A 的定理 A.4.1.

类似地，一个实数集 F 称为下有界（bounded below），它的下界（lower bound）为 v，如果对任意 $x \in F$，有 $v \leqslant x$，并且称 v 为 F 的下确界（infimum），记作 $v = \inf F$，当且仅当对 F 的任意一个下界 t，有 $t \leqslant v$．每一个下有界的非空集 F 有下确界，即（非空且上有界的）F 的下界中的上确界.

两个实数的最大值和最小值定义为：若 $x \leqslant y$，则 $\min(x, y) = x$，$\max(x, y) = y$；否则有 $\min(x, y) = y$，$\max(x, y) = x$.

对任意的实数 $a \leqslant b$．令 $[a, b] := \{x \in \mathbb{R}, a \leqslant x \leqslant b\}$.

对于任意的两个集合 X 和 Y，它们的笛卡儿积（Cartesian product），记作 $X \times Y$，定义为由 X 中的 x 和 Y 中的 y 构成的所有有序对 $\langle x, y \rangle$ 的集合．笛卡儿积的一个典型的例子是 $\mathbb{R} \times \mathbb{R}$，也可以写作 $\mathbb{R}^2$（读作 r-2，而不是 r-平方），称为平面（plane）.

习题

1. 设 $A:=\{3,4,5\}$，$B:=\{5,6,7\}$. 计算：(a)$A\cup B$. (b)$A\cap B$. (c)$A\setminus B$. (d)$A\Delta B$.
2. 证明 $\varnothing\neq\{\varnothing\}$，$\{\varnothing\}\neq\{\{\varnothing\}\}$.
3. 下列三组集合哪些是相等的？
   (a)$\{\{2,3\},\{4\}\}$. (b)$\{\{4\},\{2,3\}\}$. (c)$\{\{4\},\{3,2\}\}$.
4. 下列哪些是函数，为什么？
   (a)$\{\langle 1,2\rangle,\langle 2,3\rangle,\langle 3,1\rangle\}$.
   (b)$\{\langle 1,2\rangle,\langle 2,3\rangle,\langle 2,1\rangle\}$.
   (c)$\{\langle 2,1\rangle,\langle 3,1\rangle,\langle 1,2\rangle\}$.
   (d)$\{\langle x,y\rangle\in\mathbb{R}^2: x=y^2\}$.
   (e)$\{\langle x,y\rangle\in\mathbb{R}^2: y=x^2\}$.

8

5. 对于任意关系 V(即有序对所组成的任意集合)，定义 V 的定义域为$\{x$：对某些 y，$\langle x,y\rangle\in V\}$，$V$ 的值域为$\{y$：对某些 x，$\langle x,y\rangle\in V\}$. 求上题中各个关系的定义域和值域.(它们是否是函数?)
6. 设 $A_{1j}:=\mathbb{R}\times[j-1,j]$，$A_{2j}:=[j-1,j]\times\mathbb{R}$，其中 $j=1,2$. 设 $B:=\bigcup_{m=1}^{2}\bigcap_{n=1}^{2}A_{mn}$，$C:=\bigcap_{n=1}^{2}\bigcup_{m=1}^{2}A_{mn}$，则下面关系哪一个是正确的：$B\subset C$ 还是 $C\subset B$？为什么？
7. 令 $f(x):=\sin x$，$x\in\mathbb{R}$. 对于$\mathbb{R}$ 的下列子集，f 映射到哪一个，f 映上到哪一个？
   (a)$[-2,2]$. (b)$[0,1]$. (c)$[-1,1]$. (d)$[-\pi,\pi]$.
8. 如果将习题 7 中的弧度转换为度，会有什么结果？
9. 在下面的集合中，哪些包含于其他的集合中？
   $$A:=\{3,4,5\};\ B:=\{\{3,4\},5\};$$
   $$C:=\{5,4\};\ D:=\{\{4,5\}\}.$$
   假定有明显的关系成立，例如 $4=\{3,5\}$. 更具体一点，假定对于任意两个集合 x 和 y，下列三种关系至多有一个成立：$x\in y$，$x=y$，或 $y\in x$，并且每个非负整数 k 是一个有 k 个元素的集合. 请解释上述的关系中为何包含，为何不包含. 例如，如果 $\{\{6,7\},\{5\}\}\subset\{3,4\}$，则由延伸性，$\{6,7\}=3$ 或 4，但是 $\{6,7\}$ 只有两个元素，而不是三个或四个.
10. 令 $I:=[0,1]$，计算 $\bigcup_{x\in I}[x,2]$，$\bigcap_{x\in I}[x,2]$.
11. "半闭直线"是实数集$\mathbb{R}$ 的子集，具有形式$\{x\in\mathbb{R}: x\leqslant b\}$或$\{x\in\mathbb{R}: x\geqslant b\}$，其中的 b 是实数. $\mathbb{R}$ 上的 n 次多项式(polynomial of degree n)是一个函数
   $$x\mapsto a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0\quad 且\quad a_n\neq 0.$$
   证明：当 n 为奇数时，任意 $n\geqslant 1$ 次多项式的值域是$\mathbb{R}$；而当 n 是偶数时，其值域是半闭直线.[提示：证明当 $|x|$ 充分大时，多项式与它的首项 a_nx^n 的符号相同，并且它的绝对值趋于∞. 对连续函数(如多项式)使用介值定理(参见习题 2.2.14(d)).]
12. $\mathbb{R}^2$ 上的多项式是具有如下形式的函数：$\langle x,y\rangle\mapsto\sum_{0\leqslant i\leqslant k,0\leqslant j\leqslant k}a_{ij}x^iy^j$. 证明：$\mathbb{R}^2$ 上非常数多项式的值域或者是$\mathbb{R}$，或者是半闭直线，或者是半开直线$(b,\infty):=\{x\in\mathbb{R}: x>b\}$或$(-\infty,b):=\{x\in\mathbb{R}: x<b\}$，这里每一个开的或闭的半直线是某个多项式的值域.[提示：对一半开直线，讨论多项式 $x^2+(xy-1)^2$ 的值域.]

1.2 关系和序

一个关系(relation)就是有序对的任意集合. 对于任意的关系 E，逆关系定义为 $E^{-1}:=\{\langle y,x\rangle:$ 9 $\langle x,y\rangle\in E\}$. 因此，函数是一种特殊的关系，它的逆 f^{-1}不必是一个函数. 事实上，一个函数称为

1－1的或一对一的(one-to-one)，当且仅当f^{-1}也是一个函数．给定关系E，经常把$\langle x, y\rangle \in E$表示为$xEy$(注意：这种记号不用于函数而用于其他关系，我们将在后面解释)．给定一个集合X，关系$E\subset X\times X$称为在X上是自反的(reflexive)，当且仅当对所有的$x\in X$，有xEx. 关系E称为对称的(symmetric)，当且仅当$E=E^{-1}$. 称关系E是传递的(transitive)，当且仅当若xEy，yEz，则有xEz. 一个具有传递关系的例子是序，例如$x\leqslant y$.

一个关系$E\subset X\times X$称为等价关系(equivalence relation)，当且仅当它在X上是自反的、对称的和传递的．等价关系的一个例子是等式．一般地，一个等价关系相当于一个等式．满足等价关系的两个元素x和y，在某种意义下是相等的．例如，两个整数m和n称为在模p的意义下相等，当且仅当$m-n$能被p整除．在模p的意义下相等是一种等价关系．或者如果f是一个函数，我们可以通过xE_fy定义等价关系E_f，当且仅当$f(x)=f(y)$.

给定一个等价关系E，一个等价类(equivalence class)是具有如下形式的集合：对任意$x\in X$，$\{y\in X: yEx\}$. 从等价关系的定义推出，两个等价类或者不相交或者相等．设$f(x):=\{y\in X: yEx\}$，则f是一个函数且xEy，当且仅当$f(x)=f(y)$，因此$E=E_f$，并且每一个等价关系可以写成E_f形式.

一个关系E称为是反对称的(antisymmetric)，当且仅当若xEy且yEx，则$x=y$. 给定一个集合X，一个偏序(partial ordering)是一个传递的、反对称的关系$E\subset X\times X$，那么$\langle X, E\rangle$称为偏序集(partially ordered set)．例如，对于任意一个集合Y，令$X=2^Y$(Y的所有子集的集合)，则对于通常的包含关系$\subset$，$\langle 2^Y, \subset\rangle$构成了一个偏序集．(注：许多人认为偏序也是自反的，这里定义的偏序不但包括$\leqslant$，而且还有$<$.)一个偏序称为是严格的(strict)，如果xEx对于任意的x都不成立．因此"严格的"和"自反的"是相对的．对于任意的偏序集E，由$x\leqslant y$定义关系$\leqslant$，当且仅当(xEy或者$x=y$)，则$\leqslant$是一个自反偏序．由$x<y$定义关系$<$，当且仅当(xEy且$x\neq y$)，则$<$是一个严格偏序．例如，在实数之间通常的关系$\leqslant$和$<$就是上面所定义的偏序和严格偏序．$F=E\cup D$和$E=F\setminus D$定义了集合X的严格偏序E和自反偏序F之间的一个一一对应，这里D是"对角的"，$D:=\{\langle x, x\rangle : x\in X\}$. 从现在开始所考虑的偏序集或者是自反的，通常记作$\leqslant$或$\geqslant$；或者是严格的，记作$<$(或$>$)．这里，"$<$"通常读作"小于"，等等.

两个偏序集$\langle X, E\rangle$和$\langle Y, G\rangle$称为是有序同构的(order-isomorphic)，当且仅当存在一个从X映射到Y的一一对应(1－1且映上)的函数f，使得对任意的x，$u\in X$，有uEx，当且仅当$f(u)Gf(x)$. 那么，f称为有序同构(order-isomorphism)．例如，区间$[0, 1]$和$[0, 2]$，具有通常的实数之间的有序关系，通过函数$f(x)=2x$，它们是有序同构的．但是区间$[0, 1]$不与$\mathbb{R}$有序同构，它没有最小元素.

10

从现在开始，有序对$\langle x, y\rangle$一般记作(x, y)(这当然与无序对$\{x, y\}$是不一样的).

集合X的线性序(linear ordering)E是X的偏序E，若$\forall x$，$y\in X$，有xEy，yEx，或者$x=y$，则称$\langle X, E\rangle$为线性序集(linearly ordered set)．线性序集的一个经典例子是实直线$\mathbb{R}$，满足通常的有序关系．实际上$(\mathbb{R}, <)$、$(\mathbb{R}, \leqslant)$、$(\mathbb{R}, >)$和$(\mathbb{R}, \geqslant)$都是线性序集.

如果$\langle X, E\rangle$是任意一个偏序集，A是X的任意一个子集，则$\{\langle x, y\rangle \in E: x\in A, y\in A\}$也是$A$上的一个偏序，记作$E_A$. 对于大多数的序，就像实数上的序那样，子集的序与整个集合上的序用同一记号来表示．如果(X, E)是线性序的且$A\subset X$，则直接从定义知，(A, E_A)也是线性序的.

设W是一个具有自反线性序$\leqslant$的集合．W称为是良序的(well-ordered)，当且仅当对于W的任意非空集合A，在A中存在最小元素x，使得对A中任意的元素y，有$x\leqslant y$. 相应的严格线性序$<$也

是一个良序．如果 X 是一个有限集，则 X 中的任意一个线性序都是良序．虽然区间$[0, 1]$有最小元素0，但它不是良序的，因为它有没有最小元素的子集，比如$\{x: 0<x\leqslant 1\}$.

在良序集的证明中可以用数学归纳法．假设$(X, <)$是良序集，我们想要证明 X 中的所有元素都具有某些性质．如若不然，则必有一个最小的元素不具有该性质．因此只需证明对每个 $x\in X$，若对所有 $y<x$，该性质都成立，则对 x 它也成立．这种“归纳原理”将在1.3节详细讨论.

习题

1. 对任意偏序 E，证明 E^{-1} 也是偏序．
2. 对两个偏序集$\langle A, \leqslant\rangle$和$\langle B, \leqslant\rangle$，笛卡儿积 $A\times B$ 的*字典序*(lexicographical ordering)如下定义：$\langle a, b\rangle\leqslant\langle c, d\rangle$，当且仅当 $a<c$ 或者$(a=c, b\leqslant d)$．（例如，如果 A 和 B 两者都是具有通常次序的字母，则我们有两个字母的单词或者字母串的字典序．）如果 A 与 B 上的序是线性的，证明：$A\times B$ 上的字典序也是线性的．如果通过给定的关系，A 和 B 是良序的，证明：$A\times B$ 也是良序的．

11

3. 设$\langle a, b\rangle\leqslant\langle c, d\rangle$，当且仅当 $a\leqslant c$ 且 $b\leqslant d$，证明：这是一个偏序．如果 A 和 B 都包含不止一个元素，证明：这个序不是 $A\times B$ 上的线性序．
4. 在$\mathbb{R}^2$上，设$\langle x, y\rangle E\langle u, v\rangle$当且仅当 $x+y=u+v$，$\langle x, y\rangle F\langle u, v\rangle$当且仅当 $x+y\leqslant u+v$，$\langle x, y\rangle G\langle u, v\rangle$当且仅当 $x+u\leqslant y+v$．则 E，F 和 G 中哪一个是等价关系，哪一个是偏序或者线性序？为什么？
5. 对实数列$\{x_n\}$，令$\{x_n\}E\{y_n\}$当且仅当$\lim\limits_{n\to\infty}x_n-y_n=0$，而$\{x_n\}F\{y_n\}$当且仅当$\lim\limits_{n\to\infty}x_n-y_n=1$，则 E 和 F 中，哪一个是等价关系和(或)偏序？为什么？
6. 对于同一集合 X 上的任意两个关系 E 和 F，用 xGz 定义一个关系 $G:=E\circ F$，当且仅当对某个 y，xEy 和 yFz．对下面的每一个性质：(a)自反的，(b)对称的，(c)传递的．如果 E 和 F 都具有该性质，证明：G 也具有该性质，若不具有则举出反例．
7. 根据习题6，对如下性质回答同一问题．(a)反对称的，(b)等价关系，(c)函数．

*1.3 超限归纳和递归

数学归纳法是众所周知的并且是在证明与非负整数有关的命题时常用的方法．如果 $F(n)$ 是我们需要证明的对所有 $n\in\mathbb{N}$ 成立的命题，而直接的证明又不那样容易，那么先证 $F(0)$ 成立，再假定 $F(n)$ 成立，然后证明 $F(n+1)$ 也成立，这种方法往往是非常有帮助的．或者，也可以假设 $F(0)$，$F(1)$，…，$F(n)$ 都成立，然后证明 $F(n+1)$ 也成立．更一般地，设$(X, <)$是任意一个偏序集，子集 $Y\subset X$ 称为*归纳的*(inductive)，如果对每一个 $x\in X$，使得 $y\in Y$，对所有 $y\in X$，使得 $y<x$，就有 $x\in Y$．假如 X 有一个最小元素 x，则不存在 $y<x$，因此，x 一定属于 X 的任意一个归纳子集 Y．在一般的归纳法中，设 Y 是所有使得 $F(n)$ 成立的 n 的集合．证明 Y 是归纳的就证明了 $Y=\mathbb{N}$，于是 $F(n)$ 对所有的自然数 n 都成立．在实数集$\mathbb{R}$中，集合$(-\infty, 0)$是归纳的，但它不是所有的$\mathbb{R}$．集合$\mathbb{N}$是一个良序集，但$\mathbb{R}$不是，因为集合$\{x\in\mathbb{R}: x>1\}$没有最小元素．良序集的主要优点之一是它们允许如下对归纳法的扩展.

12

1.3.1 归纳原理 *设 X 是任意关于关系 $<$ 的良序集，Y 是 X 的任意归纳子集，则 $Y=X$.*

证明 如果 $X\setminus Y=\varnothing$，则结论显然成立.

若不然，设 y 是 $X\setminus Y$ 的最小元素，则对所有使得 $x<y$ 的 x，都有 $x\in Y$(如果 y 不是 X 的最小元素，显然成立)，因此，$y\in Y$，产生矛盾. □

对于任意线性序集$(X, \leqslant)$，*初始段*(initial segment) $Y\subset X$ 是满足如下条件的集合：当 $y\in Y$，$x<y$ 时，有 $x\in Y$．如果$(X, \leqslant)$是有通常序的实直线，Y 是初始段，则要么对某些 y 有 $Y=X$ 成立，

要么 $Y=\{x: x<y\}$，或者 $Y=\{x: x\leq y\}$ 成立.

在通常的数学归纳法中，集合 $(X,<)$ 与非负整数的集合 $\mathbb{N}$ 是有序同构的，或者与 $\mathbb{N}$ 的初始段(有限整数)关于通常序是有序同构的. 超限归纳法是指对具有比较复杂的良序 $(x,<)$ 的归纳法. 例如“双归纳法”. 为了证明命题 $F(m,n)$ 对所有的非负整数 m 和 n 都成立，首先证明 $F(0,0)$ 成立. 在证明 $F(m,n)$ 时，我们可以先假设 $F(j,k)$ 对于所有的 $j<m$ 和 $k\in\mathbb{N}$，及对 $j=m$ 和 $k<n$ 都成立.(在这种情况下，良序就是 1.2 节习题 2 中提到的字典序.)$\mathbb{N}\times\mathbb{N}$ 的其他良序也是有用的. 在集合论中，我们更关心的是一般的良序集而不仅是序列，比如 $\mathbb{R}$ 的良序，尽管这些没有本质意义上的差别.(一般集合的良序，特别是 $\mathbb{R}$ 的良序，依赖于选择公理，我们将在 1.5 节讨论.)

下面将给出数学中的另外一种重要的方法——递归. 在它的传统形式中，函数 f 可以由初始 $f(0)$，然后用 $f(n-1)$ 以及对 $k<n$ 的 $f(k)$ 其他可能的值确定 $f(n)$ 来定义. 这种递归定义可以推广到良序集上. 对于 $A\subset \operatorname{dom} f$ 和任意函数 f，函数 f 在 A 上的限制定义为 $f\upharpoonright A:=\{\langle x, f(x)\rangle: x\in A\}$.

1.3.2 递归原理 *设 $(X,<)$ 是良序集，Y 是任意集合. 对任意 $x\in X$，设 $I(x):=\{u\in X: u<x\}$. 设函数 g 的定义域是所有满足如下条件的 j 的集合：对某些 $x\in X$，j 是从 $I(x)$ 映射到 Y 的函数，并且 $\operatorname{ran} g\subset Y$，则存在唯一的从 X 映射到 Y 的函数 f，使得对每个 $x\in X$，$f(x)=g(f\upharpoonright I(x))$.*

注：*如果 b 是 X 中的最小元素，要想定义 $f(b)=c$，则只要让 $g(\varnothing)=c$，$I(b)=\varnothing$ 即可.*

证明 如果 $X=\varnothing$，那么 $f=\varnothing$，结论成立. 设 X 是非空集合，且 b 是它的最小元素. 对每个 $x\in X$，令 $J(x):=\{u\in X: u\leq x\}$，令 T 是所有满足如下条件的 $x\in X$ 的集合：在 $J(x)$ 上有一个函数 f，满足对所有 $u\in J(x)$，$f(u)=g(f\upharpoonright I(u))$. 下面证明：如果这样的函数 f 存在，则它是唯一的. 设 h 是另外一个这样的函数，则 $h(b)=g(\varnothing)=f(b)$. 由归纳原理(1.3.1)可得，对每个 $u\in J(x)$，$h(u)=g(f\upharpoonright I(u))=f(u)$. 因此 f 是唯一的. 如果对 T 中的某个 u，$x<u$，f 如上是定义在 $J(u)$ 上的函数，则 $f\upharpoonright J(x)$ 有所期望的性质，并且由唯一性知，对 $J(x)$ 来说，它就是 f. 因此，T 是 X 的一个初始段. 对所有的 $x\in T$，所有这样的函数 f 的并是一个确定的函数，我们把它也称为 f. 如果 $T\neq X$，设 u 是 $X\setminus T$ 的最小元素. 那么 $T=I(u)$，且 $f\cup\{\langle u, g(f)\rangle\}$ 是 $J(u)$ 上的具有所期望性质的函数，因此 $u\in T$，矛盾. 所以函数 f 存在，因为它在每一个 $J(x)$ 上都是唯一的，所以它是唯一的. □

[13]

对笛卡儿积 $A\times B$ 上的任意函数 f，我们通常记作 $f(a,b)$ 而不是 $f(\langle a,b\rangle)$. 在非负整数上传统的递归阐述如下.

1.3.3 推论(简单递归) *设 Y 是任意集合，$c\in Y$，h 是从 $\mathbb{N}\times Y$ 到 Y 的任意函数，则存在从 $\mathbb{N}$ 到 Y 的函数 f，使得 $f(0)=c$ 且对每个 $n\in\mathbb{N}$，$f(n+1)=h(n,f(n))$.*

证明 为了应用递归原理 1.3.2，设 $g(\varnothing)=c$，令 j 是从某个非空集合 $I(n)$ 映射到 Y 的任意函数.(注意：$I(n)$ 是空集当且仅当 $n=0$). 则 $n-1$ 是 $I(n)$ 中最大的数，设 $g(j)=h(n-1, j(n-1))$，则函数 g 由这样的函数 j 所完全确定，应用递归原理 1.3.2 得到函数 f，此时 $f(0)=c$，并且对任意 $n\in\mathbb{N}$，$f(n+1)=g(f\upharpoonright I(n+1))=h(n,f(n))$. □

例：设 t 是 $\mathbb{N}$ 上的实值函数，令

$$f(n)=\sum_{j=0}^{n} t(j).$$

为了利用简单递归 1.3.3 得到 f，对于任意的 $n\in\mathbb{N}$，$y\in\mathbb{R}$，令 $c=t(0)$ 和 $h(n,y)=t(n+1)+y$. 给出 t 的算法程序后，根据递归算法就可以写出计算 f 的计算机程序，从而在某种意义下把求和归约为简单的加法.

例：一般递归原理1.3.2可以用来定义一个函数f，使得对$n=1, 2, \cdots$，$f(n)$是第n个素数：$f(1)=2$，$f(2)=3$，$f(3)=5$，$f(4)=7$，$f(5)=11$，等等. 在空函数上，函数g定义为2，因此$f(1)=2$. 在$J(n)=\{1, 2, \cdots, n\}=I(n+1)$上给定$j$，令$g(j)$是使得$k>j(n)$且$k$不能被$j(1), j(2), \cdots, j(n)$整除的最小$k$值. 因为有无限多个素数，所以这样的$k$是存在的. 因此$f(1)<f(2)<\cdots$. 在构造函数$f$时，$g$只用于具有性质$j(1)<j(2)\cdots<j(n)$的函数$j$(因为$j(i)=f(i)$). 注意到，在$f(n+1)=g(j)$的定义中使用的是所有的$j(1), \cdots, j(n)$的值而不仅仅是$j(n)$的值.

14

习题

1. 为了用简单递归法定义阶乘函数$f(n)=n!$，如何选择推论1.3.3中的c和h? 为使h达到这个目的，对哪些n和x(如果任意)，$h(n, x)$的值是唯一确定的?
2. 设$(X, <)$是一个良序集，$>$与通常的定义一样，即$x>y$意味着$y<x$，设$(X, >)$也是一个良序集，且f是从$\mathbb{N}$映上到X的函数，证明：f不是一对一的.［提示：如果是，则在$\mathbb{N}$上通过递归定义函数h，使得$h(0)$对$<$是X的最小元素，$h(1)$是次小的元素，等等，则h的值域中没有最大元素.］
3. 给定一个偏序集$(X, \leqslant)$，子集$A\subset X$，元素$x\in A$称为A的极小(minimal)元素，当且仅当不存在$y\in A$，使得$y<x$. $(X, \leqslant)$称为极小序(min-ordered)当且仅当它的每一个非空子集有一个极小元素.(注意：任意良序集都是极小序集.)

   (a)证明：将归纳原理1.3.1中的良序换为极小序，原理仍然成立.

   (b)同样递归原理1.3.2也成立.
4. 对于习题3，在$\mathbb{N}\times\mathbb{N}$上由$(i, k)\leqslant(m, n)$定义序当且仅当$i\leqslant m$，$k\leqslant n$，

   (a) 证明：$(\mathbb{N}\times\mathbb{N}, \leqslant)$是极小序.

   (b)设$A:=\{(m, n)\in\mathbb{N}\times\mathbb{N}: m+n\geqslant 4\}$，则$A$的极小元素是什么? 证明：$(\mathbb{N}\times\mathbb{N}, \leqslant)$不是良序.

   注：在集合论基础上建立数学时，我们已经定义了$\mathbb{N}$，通过$\mathbb{N}\times\mathbb{N}$上的字典序或者如上所定义的极小序的递归，可以给出加法和乘法的定义(参见附录A.3及其参考文献).
5. 再一次对习题3，设$(X, \leqslant)$是偏序集，其中X的唯一归纳子集是X(像归纳原理中那样)，证明：$(X, \leqslant)$一定是极小序.

15

6. 设$(X, \leqslant)$是一个偏序集，在其上定义一个函数f，使得对每一个$x\in X$，有$f(x)<x$，利用递归原理证明：存在从$\mathbb{N}$映射到X的函数g，使得对所有的n，$g(n+1)<g(n)$.

1.4 势

两个集合X和Y有相同的势(same cardinality)，当且仅当它们之间存在一个从X映射到Y的一一对应的函数. 我们说这两个集合的"元素个数相同"，但是对于无限集来说，元素的"个数"指的是什么呢? 称一个集合X为有限(finite)集，当且仅当它和某个$n\in\mathbb{N}$有相同的势(这里的n代表集合$\{0, 1, \cdots, n-1\}$，它有n个元素). 否则，就称X是无限集. 例如，$\mathbb{N}$是无限集. 称X为可数(countable)集，当且仅当存在从$\mathbb{N}$映上到X的函数f，如果X还是无限的，则称X是可数无限(countably infinite)集. 一个集合是不可数(uncountable)集，当且仅当它不是可数的.

例如，$\mathbb{N}$与$\mathbb{N}\times\mathbb{N}$有相同的势，因为存在函数$f(m, n):=2^m(2n+1)-1$，它是一个从$\mathbb{N}\times\mathbb{N}$映射到$\mathbb{N}$的一一对应的函数.

称X比Y的势小(smaller cardinality)，当且仅当存在一个从X映射到Y的$1-1$的函数，但是不存在从X映上到Y的函数. 下面的事实说明这样的定义是一致的.

1.4.1 等价定理 设A，B是两个集合，f是从A映射到B的$1-1$的函数，g是从B映射到A的$1-1$

的函数，则集合 A 和 B 有相同的势.

证明 对任意的函数 j 和集合 X，令 $j[X]:=\{j(x): x\in X\}=\mathrm{ran}(j\upharpoonright X)$. 对任意 $X\subset A$，令 $F(X):=A\setminus g[B\setminus f[X]]$，对任意使得 $X\subset U\subset A$ 的 U，有 $F(X)\subset F(U)$，因为 X 越大，$B\setminus F[X]$越小，进而 $F(X)$越大. 令 $W:=\{X\subset A: X\subset F(X)\}$，$C:=\cup W$. 对任意 $u\in C$，对某个 $X\in W$，则有 $u\in X$，所以有 $u\in X\subset F(X)\subset F(C)$，因此 $C\subset F(C)$，$F(C)\subset F(F(C))$，所以 $F(C)\in W$，而由 C 的定义知，$F(C)\subset C$，因此 $F(C)=C$. 由 F 的定义，g 是从 $B\setminus f[C]$映射到 $A\setminus F(C)=A\setminus C$ 的一一对应函数，在任何情况下，f都是从 C 映射到 $f[C]$的一一对应函数，设

$$h(x):=\begin{cases}f(x), & 若\ x\in C\\ g^{-1}(x), & 若\ x\in A\setminus C,\end{cases}$$

则 h 是从 A 映射到 B 的一一对应函数. □

对任意有限的 $n=0, 1, 2, \cdots$，有 $n<2^n$：例如，$0<1$，$1<2$，$2<4$，$3<8$，等等. 对于有 n 个元素的有限集 X，X 的一切子集生成的集族 2^X 一共有 2^n 个元素. 对任意大的(无限)集合，2^X 大于 X 也成立.

16

1.4.2 定理 对于任意的集合 X，X 的势小于 2^X 的势.

证明 令 $f(x):=\{x\}$，则 f 是从 X 映射到 2^X 的 $1-1$ 的函数. 设 g 是一个从 X 映上到 2^X 的函数，令 $A:=\{x\in X: x\notin g(x)\}$. 那么对某些 y，$g(y)=A$. 如果 $y\in A$，则 $y\notin g(y)=A$，但是，如果$y\notin A=g(y)$，则 $y\in A$，矛盾. □

1.4.3 推论 $\mathbb{N}$ 比 $2^{\mathbb{N}}$的势小，所以 $2^{\mathbb{N}}$是不可数的.

由于下面的定理，我们将可数无限集合$\mathbb{N}$ 的所有子集所构成的集族 $2^{\mathbb{N}}$的势记为 c，或者称为连续统(continuum)的势.

1.4.4 定理 实数集$\mathbb{R}$ 和区间$[0, 1]:=\{x\in\mathbb{R}: 0\leqslant x\leqslant 1\}$的势均为 c.

证明 见习题. □

在$\mathbb{N}$ 和 $2^{\mathbb{N}}$之间是否存在势？这就是所谓的连续统问题(continuum problem). 猜想这样的势是不存在的，因此每一个不可数实数集与$\mathbb{R}$ 的所有集合一样有相同的势，这称为连续统假设(continuum hypothesis). 这个问题不能用集合论的公理来解决(包括选择公理，参见附录 A.2). 因此，人们可以采取连续统假设或者没有得到矛盾的否定来解决. 本书从不采取否定. 偶尔，当采取连续统假设时，我们将会指出(然而在下一节，为了承认此假设正确，选择公理也会被用到).

习题

1. 证明：对任意的可数无限集 S，存在从$\mathbb{N}$ 映射到 S 上的一一对应函数. [提示：设 h 是从$\mathbb{N}$ 映上到 S 的函数，利用 h 和递归来定义函数 f，问题就是证明确实存在应用(1.3.2)的递归.]
2. 为了理解等价定理中的条件：如果在 1.4.1 的记号中，$f(x)=y$，$g(v)=u$，我们就称 x 是 y 的原像(ancestor)，v 是 u 的原像. 我们称原像的原像还是原像. 如果 h 是从 A 映射到 B 的一一对应函数，使得对在 A 中的所有 x，要么 $h(x)=f(x)$，要么 $h(x)=g^{-1}(x)$，证明：
   (a) 如果 x 有有限的偶数个原像，比如说 $2n$，则 $h(x)=f(x)$. [提示：利用关于 n 的归纳法.]
   (b) 如果 x 有有限的奇数个原像，则 $h(x)=g^{-1}(x)$.
   (c) 如果 x 有无穷多个原像，对应于等价定理 1.4.1 的证明应如何选择 $h(x)$？

17

3. 设 X 是不可数集，而 Y 是 X 的可数子集. 证明：$X\setminus Y$ 与 X 有相同的势，假定$\mathbb{N}$ 的势比 $X\setminus Y$ 的势小.(一般来说，没有选择公理，不能证明结论，这将在 1.5 节讨论.)[提示：设 B 是 $X\setminus Y$ 的可数无穷子集，则

$B\cup Y$与B有相同的势.]

4. 证明定理1.4.4.[提示：利用等价定理1.4.1，证明：[0，1]、开区间(0，1)和$\mathbb{R}$有相同的势，对于适当的常数a，b，定义一个从$\mathbb{R}$映射到(0，1)的形为$a+b\tan^{-1}x$的一一对应函数．如果$A\subset\mathbb{N}$，令$x(A):=\sum_{n\in A}\frac{1}{2^{n+1}}$(二进制展开)，这个函数不完全是1－1的，但是利用它结合习题3证明[0，1]有势c.]
5. 设X是势不超过c的一个非空集合，Y的势为c. 证明$X\times Y$的势为c.[提示：首先考虑情况X的势为c的情况，证明$2^{\mathbb{N}}\times 2^{\mathbb{N}}$与$2^{\mathbb{N}}$有相同的势.]

1.5 选择公理及其等价形式

在欧几里得几何中，人们一直认为“平行原理”缺少像其他公理、原理那样的直观性．直到19世纪，“非欧几里得几何”的出现使得“平行原理”不再假定．同样地，在集合论中，下面的公理也不必要假定，尽管已经证明它和其他的公理是一致的，但是它在给出比平凡集还不具体、不常见、直观性不明显的结构的存在性上是很有力的工具.

选择公理(AC) 设A是集合I上的函数，且对所有$x\in I$，有$A(x)\neq\varnothing$，则存在函数f，使得对所有$x\in I$，有$f(x)\in A(x)$.

笛卡儿积$\prod_{x\in I}A(x)$定义为I上所有函数f的集合，函数f满足对一切$x\in I$，$f(x)\in A(x)$. 那么

18

AC公理说明：非空集合的笛卡儿积是非空的. 这里一般称I为指标集，f为选择函数．下面的论述似乎可“证明”选择公理：对于每个x，$A(x)$是非空的，它含有一些元素，记作$f(x)$，则f就是想要的选择函数．当I是有限集时，上面的论述是正确的，但是当I是无限集时，问题在于是否存在可以用有限多的符号表示的系统规则(因为当I是有限集时，这总能做到)来对每一个x，选择$A(x)$的具体元素．在一些结果不是很清楚的情况下，人们可以在某种意义下随机选择一个元素$f(x)$. 选择公理保证了这种“无限集情况下的随机选择”是合理的.

下面是选择公理的另一种表示形式：

AC′ *设X是任意集合，$I:=2^X\setminus\{\varnothing\}$是$X$的任意非空子集，则存在从$I$映射到$X$的函数$f$，使得对每个$A\in I$，$f(A)\in A$.*

下面将证明AC与AC′是等价的. 假定AC成立，证明AC′成立，设A是I上的恒等函数，对$\forall B\in I$，$A(B)=B$，则由AC可直接推出AC′. 另一方面，假定AC′成立，给定任意值域是非空集合的函数A，取X为A的值域的并集，那么A的每个值都是X的非空子集，所以有AC′推出AC.

选择公理还有其他的等价形式被广泛地使用．我们将证明一些著名形式的等价性. 链(chain)是一个偏序集的线性有序子集(有相同的序). 在偏序集$(X,\leqslant)$中，X的一个元素z称为极大(maximal)元素，当且仅当不存在w，使得$z<w$. 如果X不是线性序，它可能会有很多个极大元素．例如，对于平凡偏序，它的严格偏序$<$是空的，这样，每一个元素都是极大元素．X的最大值(maximum)是一个$y\in X$，使得对所有$x\in X$都有$x\leqslant y$. 最大值是一个极大元素，但是反之不然．如果没有指明序，则一般指的是包含关系.

下面给出一个关于这些概念的例子，设X是$\mathbb{R}$上长度为$b-a\leqslant 2$(这里$a\leqslant b$)的所有闭区间$[a,b]$的集族，这些区间是由包含关系构成的偏序．任意长度为2的区间是极大元素，但是不存在最大值.

下面三个命题可以证明与AC等价.

良序原理(WO) 每一个集合都能被良序化.

豪斯多夫极大原理(HMP) 对每个偏序集$(P, \leqslant)$，存在极大链$L\subset P$.

19

佐恩引理(ZL) 设$(P, \leqslant)$是偏序集，使得对每个链$L\subset P$，存在上界$y\in P$，即对一切$x\in L$，有$x\leqslant y$，则P有极大元素.

1.5.1 定理 AC、WO、HMP和ZL是等价的.

注：严格地说，如果我们假定了集合论的公理系统，例如，附录A中给出的策梅洛–弗伦克尔(Zermelo-Fraenkel)系统，那么定理1.5.1中的等价命题成立．但是证明我们将以目前为止所给出的定义为基础进行，将不明确地约定任何公理.

证明 AC会应用在AC′的形式中．AC推WO：给定一个集合X，一个选择函数c，它从X的每一个非空子集选择一个元素，以及子集$A\subset X$和在A上的序$<$，$(A, <)$称c-序(c-ordered)当且仅当$<$是A上的良序，使得对每个$x\in A$，$x=c(X\setminus\{y\in A: y<x\})$．对于任意两个$c$-序子集$(A, \leqslant)$和$(B, \lambda)$，我们将证明它们互为初始段(以相同的序)．如果不是，设x是A的最小元素，使得$\{u\in A: u\leqslant x\}$不是B的初始段．这样的x是存在的，因为初始段的任意并还是初始段．则$D:=\{u\in A: u<x\}$是A和B两者的初始段，并且在其上这两个序是一致的．因为A是c-序，$x=c(X\setminus D)$，如果$B\setminus D\neq\varnothing$，设$y$是对$\lambda$的最小元素(如图1-5所示)，则因为$B$是$c$-序，$y=c(X\setminus D)=x$，这与$x$的选择矛盾．因而$B=D$是$A$的一个初始段.

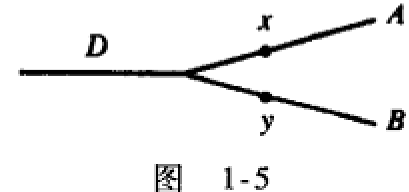


图 1-5

设C是所有c-序集的并，$\leqslant$是它们序的并，则$(C, \leqslant)$是c-序．如果$C\neq X$，令$z=c(X\setminus C)$，并且通过对所有的$x\in C$，令$x\leqslant C$(且$z\leqslant z$)来拓展序的定义，则$(C\cup\{z\}, \leqslant)$是一个c-序，出现矛盾，故X是良序.

WO⇒HMP：设$(X, \leqslant)$是偏序集，W是X的良序，t是X的关于W的最小元素(对X为空集没有问题)，通过在(X, W)上的递归(1.3.2)定义函数f，使得$f(t)=t$，并且

$$f(x)=\begin{cases}x, & \text{如果}\{x\}\cup\{f(y): yWx, y\neq x\}\text{关于}\leqslant\text{是一个链}\\ t, & \text{其他,因此 }x\text{ 关于}\leqslant\text{与某些 }y\text{、}yWx\text{ 和 }f(y)=y\text{ 不可比较.}\end{cases}$$

20

则正如我们所期望的那样，f的值域就是一个极大链.

HMP⇒ZL：设$(P, \leqslant)$是一个偏序集，它的每个链都有上界．通过HMP取一个极大链L，设z是L的一个上界．如果对某些w，$z<w$，则$L\cup\{w\}$是一个严格包含L的链，出现矛盾，因此z是极大元素.

ZL⇒AC：给定一个集合X，设F是所有选择函数的集合，其定义域是$2^X\setminus\{\varnothing\}$的子集，则通过包含关系$F$是偏序，$F$中的任意链的并还在$F$中，因此$F$有一个极大元素$f$，如果$\mathrm{dom}\, f\neq 2^X\setminus\{\varnothing\}$，那么对于非空集族$(2^X\setminus\{\varnothing\})\setminus\mathrm{dom}\, f$的任意子集$A$和对在这样的(非空)集$A$中的任意$x$，有$f\cup\{\langle A, x\rangle\}\in F$，这与$f$的极大性矛盾．因此$f$是一个所有$2^X\setminus\{\varnothing\}$上的选择函数，所以AC′成立，从而AC成立. □

习题

1. 不应用定理1.5.1证明良序原理可推出AC. 注意：尽可能大的集族$A(x)$能全部同时是良序的，因此在I上有一个函数f，使得对每一个$x\in I$，$f(x)$是$A(x)$的良序集，这显然吗？[提示：利用AC′.]
2. 证明：下面的命题和AC是等价的. 在每个偏序集中，每个链都包含在一个极大链中.
3. 在偏序集$\mathbb{N}\times\mathbb{N}$中，它的序为$(j, k)\leqslant(m, n)$当且仅当$j\leqslant m$，$k\leqslant n$. 考虑下面的序列$(n, n)$ $(n=0, 1,$

2，…）．写出包括上面所给序列的所有极大链，这些链中的任何一个都有上界吗？

4. 求一个包含尽可能少的元素的偏序集$(X,\ \leqslant)$，使得佐恩(Zorn)引理成立(每个链有一个上界)，但是没有最大值．
5. 假定 AC，证明：有限集合的任何笛卡儿积要么是有限的，要么是不可数的(不可能是无限可数的)．
6. 设 X 是所有长度为 $0\leqslant b-a\leqslant 2$ 的区间$[a,\ b]$构成的集族，包含关系为偏序．证明：X 中的每一个链都有一个上界．
7. 在 WO⇒HMP 的证明中，利用递归原理 1.3.2 时，g 应如何定义？证明得到的结论可以用来证明 WO 推 HMP.

21

8. 设 A，B 是两个集合，满足存在从 A 映上到 B 的函数 f 和从 B 映上到 A 的函数 g. 假定 AC，证明 A 和 B 有相同的势．[提示：如果对 B 中所有的 y，$h(y)\in f^{-1}(\{y\})$，则 h 是从 B 到 A 的一一对应函数．]

注释

1.1～1.3 节 集合论的注释将在附录 A 的末尾给出.

1.4 节 Georg Cantor(1874)首次证明了不可数集(所有的实数集)的存在性．在 1892 年，他证明了集合 X 的势小于其幂集 2^X 的势．1897 年，等价定理作为一个公开问题在康托尔(Cantor)的讨论班上被提出．19 岁的学生伯恩斯坦(Felix Bernstein)解决了该问题. 博雷尔(Borel)在 1898 年的书中给出伯恩斯坦的证明．同时，E. Schröder 给出另一种证明．等价定理常常称为 Schröder-Bernstein 定理．然而，Korselt 于 1911 年指出 Schröder 结论的一个错误．在 Korselt 的文章所引用的一封信中，Schröder 表示对伯恩斯坦定理深信不疑．其他注记和参考文献及这些事实的其他证明见 Fraenkel (1966，p. 70—80)．感谢 Sy Friedman 告诉了我 1.4.1 的证明．伯恩斯坦也涉足统计学(见 Frewer，1978)和遗传学，他证明了人的血型 A、B 和 O 是通过一点的三个等位基因进行遗传的(Nathan，1970；Boorman 等，1977，p. 41—43)．

1.5 节 关于选择公理及其等价性已经出版了完整的书(Jech，1973；Rubin 和 Rubin，1985)．Ernst Zermelo (1904，1908a，1908b) 和 Bertrand Russell (1906) 阐明了 AC 并利用它证明了 WO. Hausdorff(1914，p. 140—141)给出了他的极大原理(HMP)并证明它可由 WO 得出．Hausdorff (1937；1978 译著，p. 5)说“大多数序集理论”从第 1 版一直保留在以后各版．Zorn(1935)给出了著名的佐恩引理，(佐恩论述了偏序是包含关系且上界是并的情况下的 ZL)并且强调了其在代数中的应用．很容易看到它与 HMP 等价．显然佐恩是第一个提出由 ZL 或 HMP 能推出 AC 这个命题的人，

22

尽管佐恩从来没有发表过他的证明(Rubin 和 Rubin，1963，p. 11)．

参考文献

带星号的资料在次级来源中被讨论过，但在原著中并没见到.

Boorman, Kathleen E., Barbara E. Dodd, and Patrick J. Lincoln (1977). *Blood Group Serology*. 5th ed. Churchill Livingstone, Edinburgh.

*Borel, Émile (1898). *Leçons sur la théorie des fonctions*. Gauthier-Villars, Paris.

Cantor, Georg (1874). Ueber eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Journal für die reine u. angew. Math. 77*: 258–262.

——— (1892). Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der deutschen Mathematiker-Vereinigung 1*: 75–78.

Fraenkel, Abraham A. (1976). *Abstract Set Theory*. Rev. by Azriel Levy. 4th ed. North-Holland, Amsterdam.

Frewer, Magdalene (1978). Das wissenschaftliche Werk Felix Bernsteins. Diplomarbeit, Inst. f. Math. Statistik u. Wirtschaftsmath., Univ. Göttingen.

Hausdorff, Felix (1914). *Grundzüge der Mengenlehre.* Von Veit, Leipzig. 1st ed. repr. Chelsea, New York (1949). 2d ed., 1927. 3d ed., 1937, published in English as *Set Theory*, transl. J. R. Aumann et al., Chelsea, New York (1957; 2d Engl. ed., 1962; 3d Engl. ed., 1978).

Jech, Thomas (1973). *The Axiom of Choice.* North-Holland, Amsterdam.

Korselt, A. (1911). Über einen Beweis des Äquivalenzsatzes. *Math. Annalen 70*: 294–296.

Nathan, Henry (1970). Bernstein, Felix. In *Dictionary of Scientific Biography 2*: 58–59. Scribner's, New York.

Rubin, Herman, and Jean E. Rubin (1963). *Equivalents of the Axiom of Choice.* North-Holland, Amsterdam. 2d. ed. 1985.

Russell, Bertrand (1906). On some difficulties in the theory of transfinite numbers and order types. *Proc. London Math. Soc.* (Ser. 2) *4*: 29–53.

Zermelo, Ernst (1904). Beweis, dass jede Menge wohlgeordnet werden kann. *Math. Annalen 59*: 514–516.

——— (1908a). Neuer Beweis für die Möglichkeit einer Wohlordnung. *Math. Ann. 65*: 107–128.

——— (1908b). Untersuchungen über die Grundlagen der Mengenlehre I. *Math. Ann. 65*: 261–281.

Zorn, Max (1935). A remark on method in transfinite algebra. *Bull. Amer. Math. Soc. 41*: 667–670.



第2章 一般拓扑

一般拓扑主要研究与收敛概念有关的问题. 给定集合 X 中的点列 x_n, x_n 收敛到一个点 x 可以用不同的方式来定义. 主要方法之一就是用度量或者距离 d 来定义, d 是非负的并且是实值的, $x_n \to x$ 意味着 $d(x_n, x) \to 0$. 实数的通常度量是 $d(x, y) = |x-y|$. 对于通常的实数收敛性, 一个函数 f 称为是连续的, 如果当 $x_n \to x$(x 是 f 的定义域内一点)时, $f(x_n) \to f(x)$.

另一方面, 一些有用的收敛并不是通过度量来定义的: 如果定义函数序列 f_n"逐点收敛", 则 $f_n \to f$ 是指对一切 x, $f_n(x) \to f(x)$, 结果(对定义在不可数集上足够大的函数类)可能不存在一个度量 e, 使得 $f_n \to f$ 等价于 $e(f_n, f) \to 0$.

我们称集合 F 是闭的, 如果对所有的 i, 当 $x_i \in F$ 且 $x_i \to x$ 时, 有 $x \in F$. 例如, 任意闭区间 $[a, b] := \{x: a \leqslant x \leqslant b\}$ 是一个闭集. 闭集 F 与其补集 $U := X \setminus F$(它是一个开集)的性质为我们把收敛性和连续性等概念推广到非度量情况提供了最好和最可接受的方法. 这就使得各种收敛(如网格收敛和滤子收敛)比序列收敛更加一般化. 例如, 在2.2节中函数的逐点收敛就是通过"积拓扑"来处理的. 有些读者可能更喜欢在阅读2.1~2.2节之前先阅读2.3节和2.4~2.5节有关度量的内容, 这部分内容是大家比较熟悉的.

2.1 拓扑、度量和连续性

从现在起到本书最后, 我们都假定选择公理成立, 后面不再说明.

定义 给定一个集合 X, X 上的**拓扑**(topology)是由 X 的子集所构成的集族 $\mathcal{T}$(换句话说, $\mathcal{T} \subset 2^X$), 满足

24

(1) $\varnothing \in \mathcal{T}$, $X \in \mathcal{T}$.

(2) 对每一个 $U \in \mathcal{T}$ 和 $V \in \mathcal{T}$, 有 $U \cap V \in \mathcal{T}$.

(3) 对任意的 $\mathcal{U} \subset \mathcal{T}$, 有 $\cup \mathcal{U} \in \mathcal{T}$.

当一个拓扑已经选定时, 它的元素称为开集(open set). 它们的补集 $F = X \setminus U$, $U \in \mathcal{T}$ 称为闭集(closed set). 它们组成的对 $(X, \mathcal{T})$ 称为拓扑空间(topological space), X 中的元素称为点(point).

在 $\mathbb{R}$ 上有一个"普通拓扑", 对于这个拓扑, 开集的最简单的例子是开区间 $(a, b) := \{x: a < x < b\}$, 这里的 (X, Y) 是序对 $\langle X, Y \rangle$ 的记号, 这些记号在实分析中都会用到. 从上下文可以区分 X 与 Y 哪一个是实数.

在 $\mathbb{R}$ 上, 一般的开集是开区间的任意并(后面我们将看到, 任意这样的并可以写为可数并). 对于通常的实数收敛性, 闭集是开集的补集, 它和本章引言中所定义的闭集一样.

存在既不是开集也不是闭集的集合, 例如, 对任意的 a, $b \in \mathbb{R}$, 且 $a < b$, 半开区间 $[a, b) := \{x: a \leqslant x < b\}$ 或 $(a, b] := \{x: a < x \leqslant b\}$. 另一方面, 在任意的拓扑空间 X 中, 因为 $\varnothing$ 和 X 都是开集, 且它们互为补集, 所以它们也是闭集.

尽管在公理化的集合论中, 就像在附录A中那样, 所有的数学对象都可以用集合来表示. 一个点被认为是不再含有其他的元素和内部结构. 这样的点通常用小写字母来表示, 例如 x, y. 点的集合一般用大写字母来表示, 例如 A, B. 集合的集合用其他的名字, 比如"集类"或"集族", 一般用

花写的字母表示，例如$\mathcal{A}$，$\mathcal{B}$(后面在泛函分析中，由于一些特殊的需要，有时也把函数看作点).

由拓扑定义中的(2)和归纳法知，任意有限个开集的交是开集．而根据(3)知，任意有限个开集的并也是开集.

对于任意的集X，幂集2^X是拓扑，称为X上的离散拓扑(discrete topology)．对于这个拓扑，所有的集是开集并且所有的集也是闭集．如果X是一个有限集，则除了某些其他特殊的拓扑外，假定X上的拓扑是离散的.

当我们处理集合X的子集这类拓扑空间时，定义任意集合$A\subset X$的补集(complement)为$X\setminus A$. 因此，任意开集的补集是闭集.

25

定义 给定一个集合X，定义X的**伪度量**(pseudometric)为从$X\times X$映射到$\{x\in\mathbb{R}: x\geqslant 0\}$的函数$d$，满足

(1)对所有的x，$d(x, x)=0$.

(2)对所有的x，y，$d(x, y)=d(y, x)$——对称性.

(3)对所有的x，y，z，$d(x, z)\leqslant d(x, y)+d(y, z)$——三角不等式.

(4)$d(x, y)=0$当且仅当$x=y$.

则称d为一个**度量**(metric)．(X, d)称为**(伪)度量空间**.

例如，具有通常的度量$d(x, y):=|x-y|$的实数集$\mathbb{R}$就是一个度量空间．平面$\mathbb{R}^2$按照通常的欧几里得距离也构成度量空间，在这里，上述定义中的三角不等式就是所熟悉的几何中的三角形两边之和不小于第三边.

定义 拓扑$\mathcal{T}$的基是任意一个集族$\mathcal{U}\subset\mathcal{T}$，使得对每一个$V\in\mathcal{T}$，有$V=\cup\{U\in\mathcal{U}: U\subset V\}$．点$x$的**邻域**(neighborhood)是任意的集合N(开集或其他)，使得对于某些开集U，$x\in U\subset N$．点x的邻域的集族$\mathcal{N}$(开集或其他)是x的**邻域基**(neighborhood-base)，当且仅当对于某些$N\in\mathcal{N}$，x的每一个邻域V，有$x\in N\subset V$.

$\mathbb{R}$上的一般拓扑有一个由所有开区间$(a, b)(a<b)$组成的基．任意的点$x\in\mathbb{R}$的邻域基是所有形如$(x-1/n, x+1/n)(n=1, 2, \cdots)$的区间的并，例如，区间$[-1/3, 1/4]$和$(-1/8, 1/7)$都是$\mathbb{R}$上0的邻域，而$[0, 1]$不是.

空集的并还是空集(根据定义)，所以$\cup\varnothing=\cup\{\varnothing\}=\varnothing$．因此，对$V=\varnothing$，不管是否有$\varnothing\in\mathcal{U}$，$V$总是$\mathcal{U}$中包含那些集合$U$的并．如果$\mathcal{U}$是一个基，则$X=\cup\mathcal{U}$: X的每一个点都必须至少在$\mathcal{U}$中的一个集合中.

注意，一个集合U是开的，当且仅当它是它的每一个点的邻域．给定一个伪度量空间$\langle X, d\rangle$，$x\in X$，$r>0$，设$B(x, r):=\{y\in X: d(x, y)<r\}$，则称$B(x, r)$是一个以$x$为中心，以$r$为半径的(开)球.

特别地，在$\mathbb{R}$中，球$B(x, r)$是开区间$(x-r, x+r)$．反之，任意的开区间$(a, b)(a, b\in\mathbb{R}, a<b)$可以写作$B(x, r)$，其中$x=(a+b)/2$，$r=(b-a)/2$.

2.1.1 定理 对于任意的伪度量空间$\langle X, d\rangle$，所有集合$B(x, r)(x\in X, r>0)$的集族是X上的拓扑基．对于固定的x，它是这种拓扑在x的邻域基.

26

证明 设$x\in X$，$y\in X$，$r>0$，$s>0$．令$U=B(x, r)\cap B(y, s)$，假定$z\in U$，令$t:=\min(r-d(x, z), s-d(y, z))$，则$t>0$．要证明$B(z, t)\subset U$，假定$d(z, w)<t$，则由三角不等式得，$d(x, w)<d(x, z)+t<r$，同样地，$d(y, w)<s$，于是$w\in B(x, r)$，$w\in B(y, s)$，所以$B(z, t)\subset U$．因而对

U 的每一个点 z，围绕 z 的开球包含在 U 中，且 U 是所有包含 z 的开球的并.

设$\mathcal{T}$是所有开球的并的集类，于是 $U\in\mathcal{T}$. 假定 $V\in\mathcal{T}$, $W\in\mathcal{T}$, 于是 $V=\cup\mathcal{A}$, $W=\cup\mathcal{B}$其中的$\mathcal{A}$, $\mathcal{B}$都是开球的集类. 则

$$V\cap W=\cup\{A\cap B:A\in\mathcal{A},B\in\mathcal{B}\}.$$

因而有 $V\cap W\in\mathcal{T}$. 空集也在$\mathcal{T}$(作为空集的并)中，并且 X 是所有球的并. 显然，$\mathcal{T}$中任意集合的并还在$\mathcal{T}$中. 所以$\mathcal{T}$是拓扑. 同样很显然，球形成了它的一个基(并且它们确实是开集，因此结论是一致的). 假定 $x\in U\in\mathcal{T}$, 则对于某个 y 和 $r>0$, $x\in B(y, r)\subset U$. 令 $s:=r-d(x, y)$, 则 $s>0$ 且 $B(x, s)\subset U$, 所以所有圆心为 x 的球的集合是 x 的邻域基. □

定理2.1.1所给出的拓扑$\mathcal{T}$称为(伪)度量拓扑. 如果 d 是一个度量，则称$\mathcal{T}$是可度量的(metrizable)，并且由 d 度量. 在$\mathbb{R}$上用 $d(x, y):=|x-y|$ 所度量的拓扑是$\mathbb{R}$上的一般拓扑. 也就是说，拓扑有一个由开区间(a, b)所给定的基.

如果$(X, \mathcal{T})$是任意拓扑空间且 $Y\subset X$, 则容易看出$\{U\cap Y: U\in\mathcal{T}\}$也是 Y 上的一个拓扑，称为**相对拓扑**(relative topology).

设f是从集合 A 映射到集合 B 的函数，则对于集合 B 的任意子集 C, 令 $f^{-1}(C):=\{x\in A: f(x)\in C\}$, 我们称 $f^{-1}(C)$是 C 在 f 下的**逆像**(inverse image, 注意，函数 f 不需要是 1-1 的，这样 f^{-1}不一定是一个函数). 逆像对并和交的运算仍然成立：对 B 的子集的任意非空集类$\{B_i\}_{i\in I}$, $f^{-1}(\bigcup_{i\in I}B_i)=\bigcup_{i\in I}f^{-1}(B_i)$, $f^{-1}(\bigcap_{i\in I}B_i)=\bigcap_{i\in I}f^{-1}(B_i)$, 如果 I 是空集，第一个式子仍成立的，只不过，两边都等于空集；如果我们定义空间 X 上的空子集类的交与 X 是相等的($X=A$ 或 B), 则上面的第二个式子也是成立的.

注意到，定义域是$\mathbb{N}$或者所有正整数$\{n\in\mathbb{N}: n>0\}$的集合上的序列是函数，该序列 x 经常写作带有下标的形式，比如$\{x_n\}_{n\geq0}$或$\{x_n\}_{n\geq1}$, 记 $x_n:=x(n)$. 一个序列称为是在某个集合 X 中的，当且仅当它的值域包含在 X 中. 给定一个拓扑空间$(X, \mathcal{T})$, 我们称一个序列 x_n **收敛**于点 x, 记作 $x_n\to x(n\to\infty)$, 当且仅当对于点 x 的每一个邻域 A, 存在一个数 m, 使得对一切 $n\geqslant m$ 时，$x_n\in A$.

27

度量空间上的连续函数的概念可以用序列的收敛(如果 $x_n\to x$, 则 $f(x_n)\to f(x)$)或者用 ε-δ 定义来描述. 由此得出连续性(同样地，一致连续性)实际上只与拓扑有关，且具有如下的简单形式.

定义 给定两个拓扑空间$(X, \mathcal{T})$, $(Y, \mathcal{U})$, 一个从 X 映射到 Y 的函数称为是**连续的**(continuous), 当且仅当对所有的 $U\in\mathcal{U}$, $f^{-1}(U)\in\mathcal{T}$.

例：考虑从$\mathbb{R}$映射到它自身的函数f, $f(x):=x^2$, 设 $U=(a, b)$, 则逆像为

$$f^{-1}(U)=\begin{cases}\varnothing, & 若\ b\leqslant0\\(-b^{1/2},b^{1/2}), & 若\ a<0<b\\(-b^{1/2},-a^{1/2})\cup(a^{1/2},b^{1/2}), & 若\ 0\leqslant a<b.\end{cases}$$

因此在函数 f 下，一个开区间的逆像不一定是一个区间(上例中的最后一种情况，逆像是两个不相交的区间的并)，但正如在连续函数的定义中所述的那样，它一般总是一个开集. 另一方面，$f((-1, 1)):=\{f(x): -1<x<1\}=[0, 1)$, 它不是一个开区间.

如果 $n(.)$是一个从正整数映射到正整数的递增函数，即 $n(1)<n(2)<\cdots$, 则对一个序列$\{x_n\}$, 序列 $k\mapsto x_{n(k)}$称为是序列$\{x_n\}$的**子序列**(subsequence), 这里的 $n(k)$常记作 n_k.

如果 $x_n\to x$, 则对任意子序列 $k\mapsto x_{n(k)}$也收敛于 x. 设$\mathcal{T}$是一个由伪度量 d 定义的拓扑，容易看

出，对于 X 中的任意序列 $\{x_n\}$，$x_n \to x$ 当且仅当 $d(x_n, x) \to 0$ $(n \to \infty)$.

沿着一个序列收敛并不是收敛的唯一途径. 例如，称函数 f 在点 x 处是连续的一种表达方法是当 $y \to x$ 时，$f(y) \to f(x)$. 这蕴涵着对于每一个序列 $\{y_n\}$，当 $y_n \to x$ 时，有 $f(y_n) \to f(x)$，但是我们可以认为 y 是连续地趋向于 x，而不是仅仅沿着不同的序列趋向于 x. 另一方面，在某些不可度量的拓扑空间中，序列是不合适的. 例如，可能有这样的情况：对每一个 x 和每一个序列 $y_n \to x$，有 $f(y_n) \to f(x)$，但是 f 不是连续的. 有两个重要的收敛概念——"网格"和"滤子"，它们在一般的拓扑空间中所起的作用和序列在一般度量空间中所起的作用一样.

定义 一个**有向集**(directed set)是一个偏序集 $(I, \leqslant)$，使得对任意的 $i, j \in I$，有一个 $k \in I$ 且 $k \geqslant i$(即 $i \leqslant k$)和 $k \geqslant j$. 一个**网格**(net) $\{x_i\}_{i \in I}$ 是定义域为一个有向集的任意函数 x，记作 $x_i := x(i)$. [28]

设 $(X, \mathcal{T})$ 是一个拓扑空间，网格 $\{x_i\}_{i \in I}$ 收敛于 X 中的 x，记作 $x_i \to x$，当且仅当对 x 的每一个邻域 A，存在一个 $j \in I$，使得对所有的 $k \geqslant j$，$x_k \in A$.

给定一个集合 X，X 中的滤子基(filter base)是 X 的非空子集所组成的非空集类 $\mathcal{F}$，使得对任意的 $F, G \in \mathcal{F}$，对某个 $H \in \mathcal{F}$. 有 $H \subset F \cap G$. 一个滤子基 $\mathcal{F}$ 称为滤子，当且仅当若 $F \in \mathcal{F}$，且 $F \subset G \subset X$，则 $G \in \mathcal{F}$. 等价地，滤子 $\mathcal{F}$ 是 X 的非空子集所组成的非空集类，满足

(a) $F \in \mathcal{F}$, $F \subset G \subset X \Rightarrow G \in \mathcal{F}$.

(b) 如果 $F \in \mathcal{F}$，$G \in \mathcal{F}$，则 $F \cap G \in \mathcal{F}$.

例：(a)有向集的一个典型例子是具有通常序的正整数的集合. 对于它来说，网格是一个序列，因此序列是网格的一种特殊情况.

(b)另一个例子，设 I 是 $\mathbb{N}$ 的所有有限子集所组成的集合，由包含关系构成偏序. 如果 $\{x_n\}_{n \in \mathbb{N}}$ 是一个实数序列并且 $F \in I$，设 $S(F)$ 是 $x_n (n \in F)$ 的和，则 $\{S(F)\}_{F \in I}$ 是一个网格. 如果它收敛，则和 $\sum_n x_n$ 称为无条件收敛(converge unconditionally)(注意，这与绝对收敛 $\sum_n |x_n| < \infty$ 是等价的).

(c)网格的一个主要例子(这个例子比网格的概念还要早)是黎曼积分(Riemann integral). 设 a 和 b 都是实数且 $a < b$，f 是定义在区间 $[a, b]$ 上的一个实值函数，I 是所有有限序列 $a = x_0 \leqslant y_1 \leqslant x_1 \leqslant y_2 \leqslant x_2 \leqslant \cdots \leqslant y_n \leqslant x_n = b$ 的集合，这里的 n 是任意的正整数. 这样的序列记作 $u := (x_j, y_j)_{j \leqslant n}$. 假如 $v \in I$，$v = (w_i, z_j)_{j \leqslant m}$. 序定义为 $v < u$ 当且仅当 $m < n$ 并且对每一个 $j \leqslant m$，存在一个 $i \leqslant n$，使得 $x_i = w_j$.(这种关系表示区间 $[a, b]$ 的划分(partition) $\{x_0, \cdots, x_n\}$ 比划分 $\{w_0, \cdots, w_m\}$ 精细，保持 w_j 并插入一个或者多个其他的点.)容易验证这个序使得 I 成为一个有向集. 这个序不包含 y_j. 现在令 $S(f, u) := \sum_{1 \leqslant j \leqslant n} f(y_j)(x_j - x_{j-1})$，这是一个网格. 函数 f 从 a 到 b 的黎曼积分定义为这个网格的极限，当且仅当它收敛到某个实数.

如果 $\mathcal{F}$ 是任意滤子基，则 $\{G \subset X: F \subset G$，对一些 $F \in \mathcal{F}\}$ 是一个滤子 $\mathcal{G}$，$\mathcal{F}$ 称为 $\mathcal{G}$ 的一个基(base). 称滤子基 $\mathcal{F}$ 收敛于一个点 x，记作 $\mathcal{F} \to x$，当且仅当 x 的每个邻域都属于滤子. 例如，一个点 x 的所有邻域的集合是收敛于 x 的滤子. x 的所有开邻域的集合是收敛于 x 的滤子基. 设 X 是一个集合，f 是一个函数，满足 $\operatorname{dom} f \supset X$，对于每一个 $A \subset X$，$f[A] := \operatorname{ran}(f \restriction A) = \{f(x): x \in A\}$. 对于 X 中的任意滤子基 $\mathcal{F}$，令 $f[[\mathcal{F}]] := \{f[A]: A \in \mathcal{F}\}$. 注意 $f[[\mathcal{F}]]$ 也是一个滤子基. [29]

2.1.2 定理 给定拓扑空间 $(X, \mathcal{T})$，$(Y, \mathcal{U})$ 和从 X 映射到 Y 的一个函数 f，则下列命题等价(通常假定 AC 成立)：

(1) f 是连续的.

(2)对于 X 中每一个收敛的网格 $x_i \to x$，在 Y 中有 $f(x_i) \to f(x)$.

(3)对于 X 中每一个收敛的滤子基 $\mathcal{F} \to x$，在 Y 中有 $f[[\mathcal{F}]] \to f(x)$.

证明 (1)⇒(2)：设 $f(x) \in U \in \mathcal{U}$，则 $x \in f^{-1}(U)$，于是对某个 j，. 对所有的 $i > j$，有 $x_i \in f^{-1}(U)$，则 $f(x_i) \in U$，于是 $f(x_i) \to f(x)$.

(2)⇒(3)：设 $\mathcal{F} \to x$. 如果 $f[[\mathcal{F}]] \not\to f(x)$（即 $f[[\mathcal{F}]]$ 不收敛到 $f(x)$），取 $f(x) \in U \in \mathcal{U}$，使得对所有的 $A \in \mathcal{F}$，有 $f[A] \not\subset U$. 在 F 上对于 A，$B \in \mathcal{F}$. 通过"$A \leqslant B$ 当且仅当 $A \supset B$"定义偏序. 由滤子基的定义，$(\mathcal{F}, \leqslant)$ 是一个有向集. 利用 AC 通过选择定义网格，对于每一个 $A \in \mathcal{F}$，某个 $x(A) \in A$ 且 $f(x(A)) \notin U$，则网格 $x(A) \to x$，但是 $f(x(A)) \not\to f(x)$，这与(2)矛盾.

(3)⇒(1)：取任意的 $U \in \mathcal{U}$，$x \in f^{-1}(U)$，x 的所有邻域的滤子 $\mathcal{F}$ 收敛到 x. 因此 $f[[\mathcal{F}]] \to f(x)$，对于 x 的某个邻域 V，$f[V] \subset U$，于是 $V \subset f^{-1}(U)$，$f^{-1}(U) \in \mathcal{T}$. □

滤子基的另一个例子，给定区间 $[0, 1]$ 上的一个连续实函数 f，令 $t := \sup\{f(x): 0 \leqslant x \leqslant 1\}$，以递推的方式定义区间序列 I_n. 令 $I_0 := [0, 1]$，则函数 f 的上确界至少在区间 $[0, 1/2]$，$[1/2, 1]$ 中的一个上取得，设其等于 t，把这个区间记为 I_1，它是一个长度为 $\frac{1}{2}$ 的区间. 照这样下去，得到一列长度为 $1/2^n$ 的闭区间列 I_n，在它们上面 f 具有与 $[0, 1]$ 上相同的上确界 t. 设 I_{n+1} 为一个闭区间，它是 I_n 的左半区间或者右半区间，它和前面的闭区间列有相同的上确界. 则 $\{I_n\}_{n \geqslant 0}$ 是一个收敛到点 x 的滤子基，对于 x，$f(x) = t$.

一个拓扑空间 $(X, \mathcal{T})$ 称为豪斯多夫空间(Hausdorff space)，当且仅当对于 X 中两个不同的点 x，y，存在两个开集 U 和 V，使得 $x \in U$，$y \in V$. 且 $U \cap V = \varnothing$. 因此，一个伪度量空间 (X, d) 是一个豪斯多夫空间，当且仅当 d 是一个度量. 对于任意的拓扑空间 $(S, \mathcal{T})$ 和集合 $A \subset S$，A 的内部(interior)，记作 $\text{int}\,A$，定义为 $\text{int}\,A := \cup\{U \in \mathcal{T}, U \subset A\}$. 它显然是一个开集，并且是包含于 A 中的最大开集. A 的闭包(closure)，记作 $\bar{A}$，定义为 $A := \cap\{F \subset S: F \supset A$，且 F 是闭集$\}$. 容易看出，对于任意的集合 $U_i \subset S$，i 在指标集 I 中，$S \setminus (\bigcup_{i \in I} U_i) = \bigcap_{i \in I}(S \setminus U_i)$. 因为任意开集的并集是开集，任意闭集的交集是闭集，所以 $\bar{A}$ 是一个闭集，并且是包含 A 的最小闭集.

例：如果 $a < b$ 并且 A 是 (a, b)，$(a, b]$，$[a, b)$ 或 $[a, b]$ 这四个区间中的任何一个，则 $\bar{A} = [a, b]$，$\text{int}\,A = (a, b)$.

30

下面讨论闭包与收敛网格之间的关系.

2.1.3 定理 设 $(S, \mathcal{T})$ 是任意的拓扑空间，则

(a) 对于任意的 $A \subset S$，$\bar{A}$ 是使得对一切 i，$x_i \in A$，满足某些网格点 $x_i \to x$ $(x \in S)$ 的所有 x 的集合.

(b)一个集合 $F \subset S$ 是闭集，当且仅当对于 S 中的每一个网格 $x_i \to x$，对一切 i，$x_i \in F$，有 $x \in F$.

(c)一个集合 $U \subset S$，U 是开集当且仅当对于每一个 $x \in U$ 和网格 $x_i \to x$，存在某个数 j，使得对所有的 $i \geqslant j$，有 $x_i \in U$.

(d)如果 $\mathcal{T}$ 是可度量的，则上面命题(a)、(b)、(c)中的网格可以用序列 $x_n \to x$ 来替换.

证明 (a)如果 $x \notin \bar{A}$ 并且 $x_i \to x$，则对某个 i，$x_i \notin A$. 反之，如果 $x \in \bar{A}$，设 $\mathcal{F}$ 是 x 的所有邻域的滤子，则对每个 $N \in \mathcal{F}$，$N \cap A \neq \varnothing$. 利用选择公理(AC)选取 $x(N) \in N \cap A$，则网格 $x(N) \to x$（像最后的证明中那样，通过逆包含关系，邻域的集合是有向集）.

(b)注意到，F 是闭集当且仅当 $\bar{F} = F$，然后利用(a).

(c)充分性由网格收敛的定义可以得到．必要性：假定集合 B 不是开集，则对某个 $x\in B$，由(b)知，存在一个网格，使得对所有的 i，$x_i\to x$，但是 $x_i\notin B$.

(d)在(a)的证明中，我们可以取邻域的滤子基 $N=\{y: d(x, y)<1/n\}$ 来得到序列 $x_n\to x$，其余同理可证. □

对于任意的拓扑空间$(S, \mathcal{T})$，集合 $A\subset S$ 称为在 S 中是**稠密的**(dense)，当且仅当 $\bar{A}=S$. $(S, \mathcal{T})$称为是**可分的**(separable)，当且仅当它有可数个稠密子集.

例如，所有有理数的集合$\mathbb{Q}$ 在直线$\mathbb{R}$ 中是稠密的，于是$\mathbb{R}$ 是可分的(对于通常的度量).

$(S, \mathcal{T})$称为满足可数性第一公理，或者称为是**第一可数的**(first-countable)，当且仅当在每一点处有一个可数的邻域基．对于任意的伪度量空间(S, d)，拓扑是第一可数的，因为对每一点 $x\in S$，球 $B(x, 1/n):=\{y\in S: d(x, y)<1/n\}$ $(n=1, 2, \cdots)$构成了 x 的一个邻域基．(事实上，在分析中不存在其他的第一可数空间的例子．)一个拓扑空间$(S, \mathcal{T})$称为满足第二可数性公理，或者称为是**第二可数的**(second-countable)，当且仅当$\mathcal{T}$有可数基．显然，任意的第二可数空间一定是第一可数空间.

2.1.4 命题 *一个度量空间(S, d)是第二可数的当且仅当它是可分的.*

31

证明 设 A 是可数的且在 S 中稠密，$\mathcal{U}$是 A 中点 x 的所有球 $B(x, 1/n)$的构成的集合，要证明$\mathcal{U}$是一个基，设 U 是任意的开集，且 $y\in U$. 则对某个 m，$B(y, 1/m)\subset U$，取 $x\in A$ 且 $d(x, y)<1/2m$，则 $y\in B(x, 1/2m)\subset B(y, 1/m)\subset U$，因此，$U$ 是$\mathcal{U}$中所包含的元素的并集，且$\mathcal{U}$是一个基，而且还是可数的．反之，设对于这个拓扑，存在一个可数基$\mathcal{V}$，设$\mathcal{V}$由非空集合组成，由选择公理，设$\mathbb{N}$ 上的函数 f 的值域包含了$\mathcal{V}$中每一个集合的至少一个点．则这个值域是稠密的. □

习题

1. 在$\mathbb{R}^2$ 上，设 $d(\langle x, y\rangle, \langle u, v\rangle):=((x-u)^2+(y-v)^2)^{1/2}$(通常的度量)，$e(x, y):=|x-u|+|y-v|$，证明：$e$ 是一个度量，且和 d 一样度量同一个拓扑.
2. 对于任意的拓扑空间$(X, \mathcal{T})$和集 $A\subset X$. A 的**边界**(boundary)定义为$\partial A:=\bar{A}\setminus \operatorname{int} A$
   证明：A 的边界是闭的且与 $X\setminus A$ 的边界是相同的．证明：对 X 中的任意两个集合 A，B，有$\partial(A\cup B)\subset\partial A\cup\partial B$，对$\partial(A\cup B)\neq\partial A\cup\partial B$ 的情况举出一个例子.
3. 设(X, d)与(Y, e)是两个伪度量空间，分别具有由 d 和 e 所度量的的拓扑$\mathcal{T}_d$和$\mathcal{T}_e$. 设函数 f 是一个从 X 映射到 Y 的函数，证明下列命题等价：
   (a) f 是连续的，即对所有的 $U\in\mathcal{T}_e$，$f^{-1}(U)\in\mathcal{T}_d$.
   (b) f 是逐点连续的，即对每一个 $x\in X$，对于 d 的每个序列 $x_n\to x$，对于 e 则有 $f(x_n)\to f(x)$.
4. 设(S, d)是一个度量空间，X 是 S 的一个子集．我们把限制在 $X\times X$ 上的 d 仍然记作 d，证明：X 上的由 d 所度量的拓扑与在 S 上由 d 所度量的相对拓扑是相同的.
5. 证明：可分度量空间的任意子集关于其相对拓扑还是可分的．[提示：利用习题 4 和命题 2.1.4.]
6. 设$\{x_i\}_{i\in I}$是拓扑空间上的网格，定义一个滤子基$\mathcal{F}$，使得对所有的 x，$\mathcal{F}\to x$ 当且仅当 $x_i\to x$.
7. 集合 X 上的函数网格$\{f_i\}_{i\in I}$称为**逐点**(pointwise)收敛到函数 f，当且仅当对于 X 中所有的 x，有 $f_i(x)\to f(x)$，集合 A 的示性函数定义为

$$1_A(x)=\begin{cases}1, & 若\ x\in A\\ 0, & 若\ x\in X\setminus A.\end{cases}$$

如果 X 是不可数的，证明：存在一个有限集合的示性函数网格收敛到常值函数 1，但是这个网格不能由一个序列所代替.

32

8. (a)设$\mathbb{Q}$是有理数集合，证明：$I_{\mathbb{Q}}$在 0 到 1 上的黎曼积分是不确定的.（在其定义域上网格不收敛.）($\mathbb{Q}$是可数的而[0, 1]是不可数的，因此积分应该是 0，并且是它的勒贝格积分，这将在第 3 章定义）
   (b)证明：对于有限集合 $F(n)$上的示性函数序列 $1_{F(n)}$ 逐点收敛到 $1_{\mathbb{Q}}$，对每一个 n，$1_{F(n)}$的黎曼积分是 0.
9. 设 X 是一个无限集合，$\mathcal{T}$由空集和 X 中所有有限子集的补集组成. 证明：$\mathcal{T}$是一个拓扑，其中单元素集$\{x\}$是闭的，但是$\mathcal{T}$是不可度量的. [提示：不同点的序列收敛到每一个点.]
10. 设 S 是任意集合，S^{∞} 是对所有 n，使得 $x_n \in S$ 的所有序列$\{x_n\}_{n\geq 1}$所构成的集合. 设 C 是笛卡儿积$S\times S^{\infty}$的子集. $S\times S^{\infty}$ 也是对所有 $n=0, 1, \cdots$，使得 $x_n \in S$ 的所有序列$\{x_n\}_{n\geq 0}$的所构成的集合. 这样的集合 C 也可以看作是定义了一个"收敛"的状态，因此将$\{x_n\}_{n\geq 0}\in C$，记作 $x_n \to_C x_0$. 下面是一些公理：如果满足下面的条件(1)～(3)，则称 C 为 L-收敛(L-convergence).
   (1)对所有的 n，如果 $x_n = x$，则 $x_n \to_C x$.
   (2)如果 $x_n \to_C x$，则任意子序列 $x_{n(k)} \to_C x$.
   (3)如果 $x_n \to_C x$，$x_n \to_C y$，则 $x = y$.
   如果 C 还满足(4)，则称 C 为 L^*-收敛(L^*-convergence).
   (4)如果对每一个子序列 $k \mapsto x_{n(k)}$，存在一个子子列 $j \to y_j := x_{n(k(j))}$，满足 $y_j \to_C x$，则 $x_n \to_C x$.
      (a)证明：如果 T 是豪斯多夫拓扑且 $C(T)$对 T 是收敛的，则 $C(T)$为 L^*-收敛.
      (b)设 C 是任意 L-收敛. 设 $U \in T(C)$，当且仅当若 $x_n \to_C x$ 且 $x \in U$ 时，存在一个 m，使得对所有的 $n \geq m$，$x_n \in U$，证明：$T(C)$是拓扑.
      (c)设 X 是所有实数序列$\{x_n\}_{n\geq 0}$的集合，使得对某个 m，当 $n\geq m$ 时，$x_n = 0$. 若对所有的 $m = 0, 1, \cdots$，$y(m) = \{y(m)_n\}_{n\geq 0} \in X$，则称 $y(m) \to_C y(0)$. 若对某个 k，$y(m)_j = y(0)_j = 0$ 对所有的 $j \geq k$ 和所有的 m 成立，而对所有的 n，当 $m\to\infty$ 时，$y(m)_n \to y(0)_n$. 证明：$\to_C$ 是 L^*-收敛. 但是没有度量 e，使得 $y(m) \to_C y(0)$等价于 $e(y(m), y(0)) \to 0$.
11. 对任意两个实数 u 和 v，

$$\max(u,v) := \begin{cases} u, & \text{若 } u \geq v \\ v, & \text{若 } v \geq u. \end{cases}$$

   度量空间(S, d)称为超度量空间(ultrametric space)且 d 是一个超度量(ultrametric)，如果 $d(x, z) \leq \max(d(x, y), d(y, z))$，$d(x, z)$对 S 中的所有的 x，y 和 z 成立. 证明：在超度量空间中，任意开球 $B(x, r)$也是闭的.

33

2.2 紧性与积拓扑

在最优化理论中，例如，人们试图要最大化或者最小化一个函数(通常是多元函数)时，希望确切地知道在哪些条件下，最大值或者最小值存在. 就像后面定理 2.1.2 对区间[0, 1]所证明的那样，对于$\mathbb{R}$中任意的 $a\leq b$，f 是从$[a, b]$映射到$\mathbb{R}$的连续函数，则存在一个 $x\in[a, b]$，使得 $f(x) = \sup\{f(u): a\leq u\leq b\}$. 同样，存在 $y\in[a, b]$，使得$f(y) = \inf\{f(v): a\leq v\leq b\}$. 这就是我们熟悉的"闭区间上的实值连续函数有界，且能取得最大值与最小值"性质，下面将把这个结果推广到紧拓扑空间.

在一般拓扑空间之前要给出度量空间度量的紧性定义. 在度量空间中，紧性有许多等价的特性，这些将在 2.3 节中讨论. 在这些等价的特性中，下面的"海涅-博雷尔性"是用拓扑来描述的，而不是一般的度量，所以可以把这个作为一般拓扑空间上紧性的定义. 尽管这个定义没有其他的定义那样直观，但是已经证明它在数学上还是相当有用的.

定义 一个拓扑空间$(K, \mathcal{T})$称为是紧的(compact)，如果对任意的$\mathcal{U}\subset\mathcal{T}$，和 $K = \cup\mathcal{U}$，存在一个有限的$\mathcal{V}\subset\mathcal{U}$，使得 $K = \cup\mathcal{V}$.

设 X 是一个集合，A 是 X 的一个子集．如果一个集类的并包含了 A，则称这个集类为 A 的*覆盖*(cover)．如果它包含开集，则这个覆盖就称为*开覆盖*(open cover)．如果没有具体指明子集 A，则可以取 $A=X$. 所以紧性定义可以改为"每一个开覆盖有有限个子覆盖"．其中的"每一个"非常重要，因为对任意的拓扑空间来说，总是存在有有限个子覆盖的开覆盖，换句话说，存在着有限个开覆盖，事实上，开覆盖仅仅包含一个集合，因为全集 X 就是开的.

另一个例子，开区间族$(-n, n)$是$\mathbb{R}$ 的一个开覆盖，但是不存在有限个子覆盖．同样地，区间族$(1/(n+2), 1/n)(n=1, 2, \cdots)$构成了区间$(0, 1)$的一个开覆盖，没有有限个子覆盖．因而$\mathbb{R}$和$(0, 1)$都不是紧的.

一个拓扑空间 X(即$\langle X, \mathcal{T}\rangle$是拓扑空间的集合 X)的子集 K 称为是紧的，当且仅当它关于它的相对拓扑是紧的．等价地，如果对任意的$\mathcal{U}\subset\mathcal{T}$，使得$K\subset\bigcup\mathcal{U}$，存在有限个$\mathcal{V}\subset\mathcal{U}$，使得$K\subset\bigcup\mathcal{V}$，则 K 是紧的．我们知道，如果一个非空实数集合 A 有一个上界 b(即$\forall x\in A$，有 $x\leqslant b$)，则 A 有一个最小上界，称为上确界，记作 $c:=\sup A$. 也就是说，如果 c 是 A 的上确界，则对 A 的任意上界 b，都有 $c\leqslant b$(证明见 1.1 节和附录 A.4 中的定理 A.4.1)．同样，如果一个非空实数集合 D 有一个下界，则它一定有一个最大的下界，称为下确界，记作 $\inf D$. 如果集合 A 没有上界，记$\sup A:=+\infty$．如果集合 A 没有下界，记 $\inf A:=-\infty$.

34

2.2.1 定理 *任意闭区间$[a, b]$关于它的通常(相对)的拓扑是紧的.*

证明 我们只需要证明 $a=0$，$b=1$ 的情况就行了．设$\mathcal{U}$是区间$[0, 1]$的开覆盖，设 H 是$[0, 1]$中的所有 x 的集合，$[0, x]$可由$\mathcal{U}$中有限多个集合的并所覆盖．于是对某个 $V\in\mathcal{U}$，因为 $0\in V$，所以对某些 $h>0$，$[0, h]\subset H$. 如果 $H\neq[0, 1]$，令 $y:=\inf([0, 1]\setminus H)$，那么对某些 $V\in\mathcal{U}$，$y\in V$. 因此，对某个 $c>0$，$[y-c, y]\subset V$，$y-c\in H$. 取$[0, y-c]$的一个有限开子覆盖并且添加 V，给出$[0, y]$的一个开覆盖，于是有 $y\in H$. 如果 $y=1$，则结论已证明；否则，对于某个 $b>0$，$[y, y+b]\subset V$，于是$[0, y+b]\subset H$，这与 y 的选择矛盾. □

下面的两个证明是相当容易的.

2.2.2 定理 *如果$(K, \mathcal{T})$是一个紧拓扑空间，F 是 K 的闭子集，则 F 是紧的.*

证明 设$\mathcal{U}$是 F 的一个开覆盖，我们可以取$\mathcal{U}\cup\mathcal{T}$，则$\mathcal{U}\cup\{K\setminus F\}$是 K 的开覆盖．于是有一个有限子覆盖$\mathcal{V}$，那么$\mathcal{V}\setminus\{K\setminus F\}$是 F 的一个有限覆盖，且包含在$\mathcal{U}$中. □

2.2.3 定理 *如果$(K, \mathcal{T})$是紧的，且 f 是一个从 K 映上到另一个拓扑空间 L 的连续函数，则 L 是紧致的.*

证明 设$\mathcal{U}$是 L 的一个开覆盖，则$\{f^{-1}(U): U\in\mathcal{U}\}$是 K 的开覆盖，且有一个有限子覆盖$\{f^{-1}(U): U\in\mathcal{V}\}$，其中$\mathcal{V}$是有限的，则$\mathcal{V}$就是 L 的有限子覆盖. □

定理 2.2.3 的一个例子或者推论是：如果 f 是紧空间 K 上的连续函数，则 f 有界．因为$\mathbb{R}$ 中的任意紧集都是有界的(考虑区间$(-n, n)$的开覆盖).

定义 *集合 X 上的一个滤子$\mathcal{F}$称为***超滤子**(ultrafilter)，*当且仅当对于所有的 $Y\subset X$，或者 $Y\in\mathcal{F}$. 或者 $X\setminus Y\in\mathcal{F}$.*

35

最简单的超滤子是下面的形式：$\{A\subset X: x\in A\}(x\in X)$，这些称为点超滤子．非点超滤子的存在性依赖于选择公理．一些收敛于点 x 的滤子包含在所有含有 x 的集合的点超滤子中．但是，例如，$(0, 1/n)(n=1, 2, \cdots)$是$\mathbb{R}$ 中收敛于 0 的滤子基，而在基中没有集合包含 0.

下面的两个定理给出了与"在紧度量空间中每个序列都有收敛子序列"相类似的结论.

2.2.4 定理 集合 X 中的每一个滤子$\mathcal{F}$包含在某个超滤子中．$\mathcal{F}$是一个超滤子当且仅当它是包含意义下的极大值，即如果$\mathcal{F}\subset\mathcal{G}$，$\mathcal{G}$是一个滤子，则$\mathcal{F}=\mathcal{G}$.

证明 设 $Y\subset X$，如果 $F\in\mathcal{F}$，$G\in\mathcal{F}$，$F\subset Y$，且 $G\cap Y=\varnothing$，则 $F\cap G=\varnothing$，矛盾．因此，特别地，Y 中最多有一个元素及 $X\setminus Y$ 属于$\mathcal{F}$，且在 X 的滤子中，任意的超滤子都是包含意义下的极大值．或者对所有的 $G\cap\mathcal{F}$，有 $G\cap Y\neq\varnothing$，或者对所有的 $G\in\mathcal{F}$，有 $F\setminus Y\neq\varnothing$. 如果对所有的 $G\in\mathcal{F}$，$G\cap Y\neq\varnothing$，令$\mathcal{G}:=\{H\subset X$：对某个 $G\in\mathcal{F}$，$G\cap Y\subset H\}$，则显然$\mathcal{G}$是滤子，且$\mathcal{F}\subset\mathcal{G}$. 如果对某个 $G\in\mathcal{F}$，$G\cap Y=\varnothing$，则对所有的 $F\in\mathcal{F}$，有 $F\setminus Y\neq\varnothing$，定义$\mathcal{G}:=\{H\subset X$：对某个 $F\in\mathcal{F}$，$F\setminus Y\subset H\}$. 因此对于 $Y\in\mathcal{G}$，或者 $X\setminus Y\in\mathcal{G}$的滤子$\mathcal{G}$来说，总是有$\mathcal{F}\subset\mathcal{G}$，因而一个包含意义下的滤子极大值是超滤子.

下面假设$\mathcal{C}$是 X 中滤子的一个包含链且$\mathcal{U}=\cup\mathcal{C}$. 如果 $F\subset G\subset X$，$F\in\mathcal{U}$，那么对某个$\mathcal{V}\in\mathcal{C}$，$F\in\mathcal{V}$且 $G\in\mathcal{V}\subset\mathcal{U}$. 如果对某个$\mathcal{H}\in\mathcal{C}$，有 $H\in\mathcal{U}.H\in\mathcal{H}$. 或者$\mathcal{H}\subset\mathcal{V}$，或者$\mathcal{V}\subset\mathcal{H}$. 由对称性，不妨设$\mathcal{V}\subset\mathcal{H}$，则 $F\cap H\in\mathcal{H}\subset\mathcal{U}$，从而$\mathcal{U}$是滤子．因此由佐恩引理(1.5.1)，任意滤子$\mathcal{F}$被包含在某个极大的滤子中，这就是一个超滤子. □

在任何无限集中，有限子集的所有补集的集合构成了一个滤子$\mathcal{F}$. 由定理 2.2.4 知，$\mathcal{F}$包含在某个超滤子但不是点超滤子中．非点超滤子正好是包含了$\mathcal{F}$的滤子.

下面是利用超滤子来描述紧性，这就是超滤子有用的原因.

2.2.5 定理 拓扑空间$(S,\mathcal{T})$是紧的当且仅当 S 中的每一个滤子收敛.

证明 设$(S,\mathcal{T})$是紧的，$\mathcal{U}$是一个超滤子．如果$\mathcal{U}$不收敛，则对所有的 x，取开集 $U(x)$，使得 $x\in U(x)\notin\mathcal{U}$. 于是由紧性知，存在一个有限的 $F\subset S$，使得 $S=\cup\{U(x):x\in F\}$. 因为$\mathcal{U}$中的有限交还在$\mathcal{U}$中，所以有$\varnothing=\cap\{S\setminus U(x):x\in F\}\in\mathcal{U}$，矛盾．因此，每一个超滤子收敛．反之，如果$\mathcal{V}$是一个没有有限子覆盖的开覆盖，设$\mathcal{W}$是$\mathcal{V}$中所有有限并的补集所组成的集合．容易看出，$\mathcal{W}$是一个滤子基，它包含在某个滤子中，因此由定理 2.2.4 知，它包含在某个超滤子中，这个超滤子是不收敛的. □

[36]

给定一个拓扑空间$(S,\mathcal{T})$，子集类$\mathcal{U}\subset\mathcal{T}$称为是$\mathcal{T}$的子基(subbase)，当且仅当$\mathcal{U}$中集合的所有有限交集组成的集类是$\mathcal{T}$的基．例如，在实数集$\mathbb{R}$中，通常拓扑的一个子基由开半直线$(-\infty,b):=\{x:x<b\}$和$(a,\infty):=\{x:x>a\}$给出，但是不构成基．把后者中的一个和前者中的一个相交给出(a,b)，这样的区间构成了基.

2.2.6 定理 对于任意的集合 X 和由 X 的子集组成的集类$\mathcal{U}$，存在一个包含$\mathcal{U}$的最小拓扑$\mathcal{T}$，且$\mathcal{U}$是$\mathcal{T}$的子基．给定一个拓扑$\mathcal{T}$，$\mathcal{U}\subset\mathcal{T}$，$\mathcal{U}$是$\mathcal{T}$的子基当且仅当$\mathcal{T}$是包含$\mathcal{U}$的最小拓扑.

证明 设$\mathcal{B}$是$\mathcal{U}$中元素的所有有限交组成的集类，$\mathcal{B}$中某个元素是没有$\mathcal{U}$中元素的交，在这种情况下就定义该元素为 X. 设$\mathcal{T}$是$\mathcal{B}$中元素一切(任意)并集所组成的集类．下面将证明，$\mathcal{T}$是一个拓扑，$\mathcal{B}$是它的基．首先，$X\in\mathcal{B}$给出 $X\in\mathcal{T}$，并且$\varnothing\in\mathcal{T}$.

显然$\mathcal{T}$中集合的并也在$\mathcal{T}$中．于是问题变成了证明$\mathcal{T}$中任意两个集合 V 和 W 的交还在$\mathcal{T}$中．现在 V 是集类$\mathcal{V}$的并集，W 是集类$\mathcal{W}$的并集，$\mathcal{V}\cup\mathcal{W}$中的每一个集合是$\mathcal{U}$中集合的有限交．交集 $V\cap W$ 是所有交集 $A\cap B$ 的并集，其中 $A\in\mathcal{V}$，$B\in\mathcal{W}$. 但是两个有限交集的交集是有限交，因此，每个这样的集合 $A\cap B$ 在$\mathcal{B}$中．由此推出 $V\cap W\in\mathcal{T}$，因此，$\mathcal{T}$是一个拓扑．则显然$\mathcal{B}$是它的一个基，$\mathcal{U}$是子基.

由拓扑的定义，任意包含$\mathcal{U}$的拓扑必定包含$\mathcal{B}$，于是必然包含$\mathcal{T}$. 因此$\mathcal{T}$是包含$\mathcal{U}$的最小拓扑．子基$\mathcal{U}$决定基$\mathcal{B}$，进而拓扑$\mathcal{T}$是唯一的，于是$\mathcal{U}$是$\mathcal{T}$的子基当且仅当$\mathcal{T}$是包含$\mathcal{U}$的最小拓扑. □

2.2.7 推论 (a)如果$(S, \mathcal{V})$，$(X, \mathcal{T})$是拓扑空间，$\mathcal{U}$是$\mathcal{T}$的子基，f是一个从S映射到X的函数，则f是连续的当且仅当对每一个$U\in\mathcal{U}$，有$f^{-1}(U)\in\mathcal{V}$.

37

(b)设S，I是任意的集合，对每一个$i\in I$，f_i是从S映射到X_i的函数，其中$(X_i, \mathcal{T}_i)$是一个拓扑空间，则在S上存在一个最小拓扑$\mathcal{T}$，使得每一个f_i是连续的．这里$\mathcal{T}$的一个子基是由$\{f_i^{-1}(U): i\in I, U\in\mathcal{T}_i\}$给出的，并且对不同的$i$值，基是由这种集合的有限交所生成，这里的每一个$\mathcal{T}_i$都能由它的子基所代替.

证明 (a)这个结论基本上从函数的逆像$B\mapsto f^{-1}(B)$保持集合的(任意)并和交运算的事实来证明．特别地，为了证明“必要性”(充分性是显然的)，对于$\mathcal{U}$中的任意有限个集合$U_1, \cdots, U_n$，

$$f^{-1}\left(\bigcap_{1\leqslant i\leqslant n} U_i\right)=\bigcap_{1\leqslant i\leqslant n} f^{-1}(U_i)\in\mathcal{V},$$

因此，对于$\mathcal{T}$的基$\mathcal{B}$中的每个A，有$f^{-1}(A)\in\mathcal{V}$. 于是对每一个$W\in\mathcal{T}$，W是某个集类$\mathcal{W}\subset\mathcal{B}$的并．因此，

$$f^{-1}(W)=f^{-1}(\cup\mathcal{W})=\cup\{f^{-1}(B): B\in\mathcal{W}\}\in\mathcal{T},$$

这就证明了(a).

(b)由子基的定义及定理2.2.6，(b)是显然的．当我们取集合$f_i^{-1}(U_i)$的有限交为一个基时，如果对于一个i值，有超过一个的U_i，比如说$U_{ij}(j=1, \cdots, k)$，那么集合$f_i^{-1}(U_{ij})(j=1, \cdots, k)$的交等于$f_i^{-1}(U_i)$，其中$U_i$对$j=1, \cdots, k$是$U_{ij}$的交，或者如果$U_{ij}$都属于$\mathcal{T}_i$的基$\mathcal{B}_i$，那么它们的交$U_i$是包含在$\mathcal{B}_i$中的集类$\mathcal{U}_i$的并．对一个有限集合$G$中的$i$，$f_i^{-1}(U_i)$的交是$f_i^{-1}(V_i)$的所有交的并，这里对每个$i\in G$，$V_i\in\mathcal{U}_i$，于是，我们得到所描述的基. □

推论2.2.7(a)可以简化函数是连续的证明．例如，如果f是实值的，那么利用上面提到的$\mathbb{R}$中拓扑的子基，只需证明对任意的实数a，b，$f^{-1}((a, \infty))$，$f^{-1}((-\infty, b))$是开集.

对集合I中所有的$i\in I$，设$(X_i, \mathcal{T}_i)$是拓扑空间．令X是笛卡儿积$X:=\prod_{i\in I} X_i$，换句话说，是所有指标族$\{x_i\}_{i\in I}$的集合，这里对所有的i，$x_i\in X_i$. 设p_i是把X映上到第i个坐标空间X_i：$p_i(\{x_j\}_{j\in I}):=x_i$上的投影，其中$i\in I$. 在推论2.2.7(b)中令$f_i=p_i$，则给出了$X$上的拓扑$\mathcal{T}$，称为积拓扑(product topology)，它是使所有的坐标投影连续的最小拓扑.

设$\mathbb{R}^k:=\{x=(x_1, \cdots, x_k)$：对所有的$j$，$x_j\in\mathbb{R}\}$是$\mathbb{R}$的$k$次笛卡儿积，且具有积拓扑．$K$元有序组$(x_1, \cdots, x_k)$可以定义为从$\{1, 2, \cdots, k\}$映射到$\mathbb{R}$的一个函数．也可以记作$x=\{x_j\}_{1\leqslant j\leqslant k}=\{x_j\}_{j=1}^k$. $\mathbb{R}^k$上的积拓扑可以用欧几里得距离来度量(习题16). 对于任意的实数$M>0$，由定理2.2.1知，区间$[-M, M]$是紧的. $\mathbb{R}^k$中的立方体$[-M, M]^k:=\{\{x_j\}_{j=1}^k: |x_j|\leqslant M, j=1, \cdots, k\}$对于积拓扑来说是紧致的，这可以作为下面一般定理的一个特例.

38

2.2.8 定理(吉洪诺夫定理) 对集合I中的每个i，设$(K_i, \mathcal{T}_i)$是一个紧的拓扑空间，则具有积拓扑的笛卡儿积$\prod_i K_i$是紧的.

证明 设$\mathcal{U}$是$\prod_i K_i$中的超滤子，则对所有的i，$p_i[[\mathcal{U}]]$是K_i中的超滤子，因为对每一个集合$A\subset K_i$，或者$p_i^{-1}(A)$或者它的补集$p_i^{-1}(K_i\setminus A)$在$\mathcal{U}$中．于是由定理2.2.5知，$p_i[[\mathcal{U}]]$收敛到某个$x_i\in K_i$. 对于$x:=\{x_i\}_{i\in I}$的任意邻域U，由积拓扑的定义知，存在有限集$F\subset I$，$U_i\in\mathcal{T}(i\in F)$，使得$x\in\cap\{p_i^{-1}(U_i): i\in F\}\subset U$. 对每一个$i\in F$，$p_i^{-1}(U_i)\in\mathcal{U}$，因此$U\in\mathcal{U}$且$\mathcal{U}\to x$，于是每一个超滤子

都是收敛的，再一次应用定理 2.2.5 得，$\prod_i K_i$ 是紧的. □

考虑超滤子主要是为了得到最后的证明，吉洪诺夫(Tychonoff)定理的其他的证明都比较长.

在紧空间中，豪斯多夫空间有许多特殊的好性质，也是被研究得最多的.（具有相对拓扑的豪斯多夫空间的子集也是豪斯多夫空间.）下面是结合性质的一个优点.

2.2.9 命题 *豪斯多夫空间中的任意紧集 K 都是闭的.*

证明 对于任意的 $x \in K$，$y \notin K$，取开集 $U(x, y)$ 和 $V(x, y)$，满足 $x \in U(x, y)$，$y \in V(x, y)$，且 $U(x, y) \cap V(x, y) = \varnothing$. 对于每一个固定的 y，所有 $U(x, y)$ 的集合构成了 K 的具有有限子覆盖的开覆盖. 相应的有限多个 $V(x, y)$ 的交集给出了 y 的一个开邻域 $W(y)$，其中 $W(y)$ 与 K 是不相交的. 所有这样的 $W(y)$ 的并集是补集 $X \setminus K$，并且是开的. □

在任意一个集合 S 上，密着拓扑(indiscrete topology)是最小拓扑 $\{\varnothing, S\}$. S 的所有子集是紧的，但是只有 S 与 $\varnothing$ 是闭的. 这与豪斯多夫空间的一般情况相反.

设 f 是一个从 X 映射到 Y 的函数，g 是一个从 Y 映射到 Z 的函数，对所有的 $x \in X$，定义函数 $g \circ f$ 为 $(g \circ f)(x) := g(f(x))$. 于是 $g \circ f$ 是一个从 X 映射到 Z 的函数. 我们把这个函数称为 g 和 f 的复合函数(composition). 对任意的 $A \subset Z$，$(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$，于是我们有下面的定理.

39

2.2.10 定理 *如果 $(X, \mathcal{S})$，$(Y, \mathcal{T})$ 和 $(Z, \mathcal{U})$ 是拓扑空间，f 是从 X 映射到 Y 的连续函数，而 g 是从 Y 映射到 Z 的连续函数，则 $g \circ f$ 是从 X 映射到 Z 的连续函数.*

两个连续函数的复合函数的连续性很显然从网格收敛(定理 2.1.2)的连续性的表示而得到：若 $x_i \to x$，则 $f(x_i) \to f(x)$，因此 $g(f(x_i)) \to g(f(x))$.

设 $(X, \mathcal{S})$，$(Y, \mathcal{T})$ 是两个拓扑空间，从 X 到 Y 的同胚(homeomorphism)是一个从 X 到 Y 的一一对应函数 f，且 f 和 f^{-1} 都是连续的. 如果这样的函数 f 存在，则称 $(X, \mathcal{S})$ 和 $(Y, \mathcal{T})$ 是同胚的(homeomorphic). 例如，通过一个线性变换 $f(x) := a + (b - a)x$，一个有限非空的开区间 (a, b) 与 $(0, 1)$ 是同胚的. 更令人惊奇的是 $(-1, 1)$ 与 $\mathbb{R}$ 是同胚的，只需令 $f(x) := \tan(\pi x/2)$ 即可.

一般地，如果 $f \circ h$ 是连续的，且 h 是连续的，则 f 不必是连续的. 例如，h 和 $f \circ h$ 可以是常数，而 f 是任意函数. 或者设 $\mathcal{T}$ 是一个 X 上的离散拓扑 2^X，h 是一个从 X 映射到拓扑空间 Y 的函数，f 是一个从 Y 映射到另一个拓扑空间的函数，那么 h 和 $f \circ h$ 总是连续的，但 f 不必是连续的. 事实上，它可以是任意的. 然而，在下面的情况下，只要紧性和豪斯多夫同时满足，则 f 是连续的.

2.2.11 定理 *设 h 是一个从紧拓扑空间 T 映射到豪斯多夫拓扑空间 K 的连续函数，则集合 $A \subset K$ 是开的，当且仅当 $h^{-1}(A)$ 在 T 中是开的. 如果 f 是从 K 映射到另外一个拓扑空间 S 的函数，则 f 是连续函数，当且仅当 $f \circ h$ 是连续的. 如果 h 是 $1-1$ 的，则它是一个同胚.*

证明 由定理 2.2.3 知，K 是紧的，设 $h^{-1}(A)$ 是开的，于是 $T \setminus h^{-1}(A)$ 是闭的且由定理 2.2.2 知，它还是紧的. 因而由定理 2.2.3 知，$h[T \setminus h^{-1}(A)] = K \setminus A$ 是紧的，由命题 2.2.9 知，它还是闭的，于是 A 是开的. 如果 $f \circ h$ 是连续的，则对任意的开集 $U \subset S$，$(f \circ h)^{-1}(U) = h^{-1}(f^{-1}(U))$ 是开的. 于是 $f^{-1}(U)$ 是开的且 f 是连续的. 其他的结论由定义和定理 2.2.10 立即可得. □

40

集合 X 的所有子集所组成的集类 2^X（即幂集）可以看作是从 X 映射到 $\{0, 1\}$ 的示性函数的集合. 换句话说，2^X 是由 X 诱导的笛卡儿积，是 $\{0, 1\}$ 的复制. 2^X 上的积拓扑关于 $\{0, 1\}$ 上的离散拓扑是紧的.

下面的这些特殊结果将在第 12 章和第 13 章用到. 这里 $f(V) := \{f(x) : x \in V\}$ 和“ $-$ ”表示闭集.

$F_n \downarrow K$ 表示对所有的 $n \in \mathbb{N}$，$F_n \supset F_{n+1}$，且 $\bigcap_n F_n = K$.

***2.2.12 定理** 设 X 和 Y 是豪斯多夫拓扑空间，f 是一个从 X 映射到 Y 的连续函数，设 F_n 是 X 中的闭集，且当 $n \to \infty$ 时，$F_n \downarrow K$，这里的 K 是紧的. 假定有下面的两种情况之一出现：

(a)对每一个开集 $U \supset K$，存在 n，使得 $F_n \subset U$，

(b)所有的 F_n 是紧的.

则

$$f(K) = \bigcap_n f(F_n) = \bigcap_n f(F_n)^{-}.$$

证明 显然，

$$f(K) \subset \bigcap_n f(F_n) \subset \bigcap_n f(F_n)^{-}.$$

反之，假设(a)成立，取任意 $y \in \bigcap_n f(F_n)^{-}$. 假定对 K 中每个 x 有一个开邻域 V_x，使得 $y \notin f(V_x)^{-}$，则 V_x 构成了 K 的一个开覆盖，且有有限子覆盖. 在子覆盖中 V_x 的并给出了一个开集 $U \supset K$ 且 $y \notin f(U)^{-}$，这与当 n 很大时 $F_n \subset U$ 矛盾. 因此对每一个包含 x 的开集 V，取 $x \in K$，使得 $y \in f(V)$. 如果 $f(x) \neq y$，则取 $f(x)$ 的开邻域 W 和 y 的开邻域 T，使得它们不相交. 令 $V = f^{-1}(W)$，则得到一个矛盾. 因此 $f(x) = y$ 且 $y \in f(K)$，从而证明了(a).

现在只要证明(b)蕴涵(a)即可完成证明. 集列 F_n 全是紧集因为它们是紧集 F_1 的闭子集. 设开集 $U \supset K$，其中 U 是开集，则 $F_n \setminus U$ 是递减的紧集列，且交集是空集，因此对某个 n，$F_n \setminus U$ 是空的(否则，U 和 F_j 的补集将构成 F_1 的开覆盖，但是没有有限的子覆盖)，故 $F_n \subset U$. □

习题

1. 如果 S_i 是一个有离散拓扑的集列，证明：有限多个这样的空间的积拓扑也是离散的.
2. 如果有无限多个离散拓扑 S_i，每一个至少有一个点，证明：它们的积拓扑不是离散的.

41

3. 证明：可数多个可分拓扑空间的积空间关于积拓扑还是可分的.
4. 如果$(X, \mathcal{T})$，$(Y, \mathcal{U})$是拓扑空间，$\mathcal{A}$是$\mathcal{T}$的基，$\mathcal{B}$是$\mathcal{U}$的基，证明：对于 $A \in \mathcal{A}$，$B \in \mathcal{B}$，所有集合 $A \times B$ 所生成的集类是 $X \times Y$ 上积拓扑的基.
5. (a)证明：一个集合上的任意拓扑交是拓扑.
   (b)证明：对于集合 X 的子集的任意集类$\mathcal{U}$，在 X 上有一个包含$\mathcal{U}$的最小拓扑. 利用(a)(而不是子基).
6. (a)设 A_n 是所有大于 n 的整数集合，设$\mathcal{B}_n$是$\{1, \cdots, n\}$的所有子集组成的集类，$\mathcal{T}_n$是"或者在$\mathcal{B}_n$中或者具有形式 $A_n \cup B(B \in \mathcal{B}_n)$"的正整数集合组成的集类，证明$\mathcal{T}_n$是拓扑.
   (b)证明：对 $n = 1, 2, \cdots$，$\mathcal{T}_n$是一个拓扑包含链，其并不是拓扑.
   (c)对所有的 n，写出包含$\mathcal{T}_n$的最小拓扑.
7. 给定拓扑空间$(X_i, \mathcal{T}_i)$的积 $\prod_{i \in I} X_i$，具有积拓扑和有向集 J，指标在 J 中的 X 的网格通过双指标族$\{x_{ji}\}_{j \in J, i \in I}$给出. 证明：这种网格关于积拓扑收敛的充要条件是对每一个 $i \in I$，网格$\{x_{ji}\}_{j \in J}$在 X_i 中关于$\mathcal{T}_i$收敛.(由此，积拓扑有时也称为"逐点收敛"：对每个 j，有一个在 I 上的函数$i \mapsto x_{ij}$，积拓扑的收敛性等价于在每一点 $i \in I$ 的收敛性. 特别地，当 X_i 是某些空间的复制时，这种情况会出现，例如具有通常拓扑的$\mathbb{R}$. 那么 $\prod_i X_i$ 是从 I 映射到$\mathbb{R}$的所有函数的集合，通常记为$\mathbb{R}^I$).
8. 设 $I := [0, 1]$具有通常的拓扑，设 I^I 是从 I 映射到 I 的具有积拓扑的所有函数的集合.
   (a)证明：I^I 是可分的. [提示：考虑有限和函数 $\sum a_i 1_{J(i)}$，其中 a_i 是有理数，$J(i)$是具有有理数端点的

区间.]

(b)证明：I'有一个关于相对拓扑不可分的子集.

9. (a)对于任意的偏序集$(X, <)$. 集类$\{X\} \cup \{\{x: x<y\}: y \in X\} \cup \{\{x: x>z\}: z \in X\}$是$X$上的拓扑的一个子基，称为区间拓扑(interval topology). 对于实数的一般线性序，证明：区间拓扑是通常的拓扑.

42

(b)假定选择公理成立，存在一个不可数的良序集$(X, \leqslant)$. 证明：存在这样的一个集合，它恰好包含一个元素x，使得对不可数的$y \in X$的值，有$y<x$. 令$f(x)=1$且对所有其他的$y \in X$, $f(y)=0$、对于X上的区间拓扑，证明：f是不连续的，但对每一个X中的序列$u_n \to u$, $f(u_n)$收敛于$f(u)$.

10. 设f是一个定义在X上的实值有界函数：对某$M<\infty . f[X] \subset [-M, M]$. 设$\mathcal{U}$是$X$中的超滤子. 证明：$\{f[A]: A \in \mathcal{U}\}$是一个收敛的滤子基.

11. 如果函数f是无界的，习题10会出现什么情况？

12. 证明：在定理2.2.12中，如果没有“对每一个开集$U \supset K$，存在n，使得$F_n \subset U$”这一假设则该定理的结论不一定成立.[提示：设$F_n=[n, \infty)$.]

13. 对仅有两个紧集F_j的交，如果不包含在其他集合中，证明：定理2.2.12不成立.

14. 一个拓扑空间$(S, \mathcal{T})$称为是连通的(connected)，如果S不是两个不相交的非空开集的并.

(a)证明：如果S是连通的且f是一个从S映上到T的连续函数，则T也是连通的.

(b)证明：对于$\mathbb{R}$中任意的$a<b$，区间$[a, b]$是连通的. [提示：$[a, b]=U \cup V$，其中U和V是不相交的、非空的开集. 假定$c \in U$, $d \in V$，且$c<d$，令$t:=\sup(U \cap [c, d])$，则$t \in U$，或者$t \in V$，从而出现矛盾.]

(c)如果$S \subset \mathbb{R}$是连通的，且$c, d \in S$, $c<d$，证明$[c, d] \subset S$. [提示：假定$c<t<d$，且$t \notin S$，考虑$(-\infty, t) \cap S$和$(t, \infty) \cap S$.]

(d)(介值定理). 设$a, b \in \mathbb{R}$, $a<b$, f是从$[a, b]$映射到$\mathbb{R}$的连续函数，证明：f取尽$f(a)$与$f(b)$之间所有的值.[提示：应用(a)、(b)和(c).]

15. 对$\mathbb{R}^k$中的x和y，定义点积(dot product)或者内积(inner product)为$x \cdot y:=(x, y): \sum_{j=1}^{k} x_j y_j$，$x$的长度(length)定义为$|x|:=(x, x)^{\frac{1}{2}}$.

(a)(柯西不等式). 证明：对任意的$x, y \in \mathbb{R}^k$, $(x, y)^2 \leqslant |x|^2 |y|^2$. [提示：二次函数$q(t):=|x+ty|^2$未必有两个不相等的实根.]

(b)证明：对任意的$x, y \in \mathbb{R}^k$, $|x+y| \leqslant |x|+|y|$.

(c)对$x, y \in \mathbb{R}^k$，设$d(x, y):=|x-y|$，证明：d是$\mathbb{R}^k$上的度量. 它称为一般度量或者欧几里得度量.

16. 对习题15所定义的d.

(a)证明：d度量了$\mathbb{R}^k$上的积拓扑.[提示：证明任意的开球$B(x, r)$包含了开区间$(x_i-u, x_i+u)(u>0)$的积，反之亦然.]

43

(b)证明：$\mathbb{R}^k$中的任意闭集F关于d有界(即$\sup\{d(x, y): x, y \in F\}<\infty$)，且是紧集.[提示：它是一个闭区间积的子集.]

17. 一个拓扑空间$(S, \mathcal{T})$称为T_1，如果所有的单元素集$\{x\}(x \in S)$都是闭的. 设S是任意的集合.

(a)证明：空集和所有有限集合的补集所组成的集类构成了一个S上的T_1拓扑$\mathcal{T}$，其中所有的子集是紧集.

(b)如果在(a)中关于拓扑S是无限集，证明：存在非空紧子集序列$K_1 \supset K_2 \supset \cdots \supset K_n \supset \cdots$，使得$\bigcap_{n=1}^{\infty} K_n = \varnothing$.

(c)证明：(b)的情况在豪斯多夫空间不可能出现.[提示：利用命题2.2.9.]

18. 拓扑空间S上的实值函数f称为是上半连续的(upper semicontinuous)，当且仅当对于每一个$a \in \mathbb{R}$,

$f^{-1}([a, \infty))$是闭的. f称为是下半连续的(lower semicontinuous),当且仅当$-f$是上半连续的.

(a)证明:f是上半连续的,当且仅当对所有的$x \in S$,

$$f(x) \geqslant \limsup_{y \to x} f(y) := \inf\{\sup\{f(y): y \in U, y \neq x\}: x \in U \text{ 开集}\},$$

其中$\sup \varnothing := -\infty$.

(b)证明:f是连续函数,当且仅当它既是上半连续又是下半连续的.

(c)设f是紧空间S上的上半连续函数,证明:对某个$t \in S$,

$$f(t) = \sup f := \sup\{f(x): x \in S\}.$$

[提示:设$a_n \in \mathbb{R}$, $a_n \uparrow \sup f$. 考虑$f^{-1}((-\infty, a_n))$, $n=1, 2, \cdots$.]

2.3 完备度量空间和紧度量空间

空间S上的序列$\{x_n\}$关于(伪)度量d称为柯西序列(Cauchy sequence),如果$\lim_{n \to \infty} \sup_{m \geqslant n} d(x_m, x_n) = 0$. 伪度量空间$(S, d)$称为是完备的(complete),当且仅当它中的每一个柯西序列是收敛的. 拓扑空间中的一个点x称为是集合E的极限点(limit point),当且仅当x的每一个邻域包含E中除了x之外的点. 前面介绍过,对任意序列$\{x_n\}$,一个子序列是序列$k \mapsto x_{n(k)}$,这里$k \mapsto n(k)$是一个从$\mathbb{N} \setminus \{0\}$映射到自身的严格增函数.(有些作者仅仅要求当$k \to +\infty$时,$n(k) \to \infty$). 作为紧度量空间的例子,首先考虑区间$[0, 1]$, $[0, 1]$中的每个数字x都有一个十进制展开$x = 0.d_1d_2d_3\cdots$,即$x = \sum_{j \geqslant 1} d_j/10^j$. 这里每一个$d_j = d_j(x)$是一个整数并且对所有的$j$, $0 \leqslant d_j \leqslant 9$. 如果对某个$m$,当$j > m$, $d_m < 9$时,数x有一个具有$d_j = 9$的展开,那么d_j不是唯一确定的,且

$$0.d_1d_2d_3\cdots d_{m-1}d_m9999\cdots = 0.d_1d_2d_3\cdots d_{m-1}(d_m + 1)0000\cdots$$

在所有其他情形下,数字d_j都是唯一给定x.

在实际中,我们只需要十进制展开中的前几个数字. 例如,我们用$\pi = 3.14$或者$\pi = 3.1416$,很少需要知道$\pi = 3.14159265358979\cdots$. 这表明了区间$[0, 1]$中的数的一个重要性质:给定一个描述的精度(特别是给定任意的$\varepsilon > 0$),存在$[0, 1]$中的数的有限集F,使得$[0, 1]$中的每一个数x都可以用F中的一个数y来表示,并且达到要求的精度,即$|x - y| < \varepsilon$. 事实上,存在某个n,使得$\frac{1}{10^n} < \varepsilon$,因此,我们可以令$F$是具有$n$个数字的所有有限十进制展开的集合,恰好有$10^n$个这样的数字. 对于$[0, 1]$中任意的$x$,我们有$|x - 0.x_1x_2\cdots x_n| \leqslant \frac{1}{10^n} < \varepsilon$且$0.x_1x_2\cdots x_n \in F$.

上面的性质扩展到度量空间是如下的形式.

定义 一个度量空间(S, d)称为是完全有界的(totally bounded),当且仅当对每个$\varepsilon > 0$,存在有限集$F \subset S$,使得对每个$x \in S$,存在一个$y \in F$,满足$d(x, y) < \varepsilon$.

区间$[0, 1]$中实数的十进制展开的另一个便利性质是,对于$0 \sim 9$之间的整数的任意序列$x_1, x_2, \cdots$,存在一个实数$x \in [0, 1]$,使得$x = 0.x_1x_2x_3\cdots$,换句话说,特殊的柯西序列$0.x_1$, $0.x_1x_2$, $0.x_1x_2x_3$, $\cdots$, $0.x_1x_2x_3\cdots x_n$, $\cdots$实际上收敛到某一个极限值x. 区间$[0, 1]$的这个性质是度量空间完备性的一个例子(当然,不是$[0, 1]$上的所有柯西序列都具有刚才表示的特殊类型).

下面是紧度量空间的一些有用的一般性质:

2.3.1 定理 对于任意的度量空间(S, d),下面的性质是等价的(任何一个都可推出其余三个):

(Ⅰ)(S, d)是紧的,即每一个开覆盖有一个有限子覆盖.

44

(Ⅱ)(S, d)是完备的并且是完全有界的.

(Ⅲ)S 的每一个无限子集都有一个极限点.

(Ⅳ)S 的每一个点列有一个收敛的子序列.

证明　(Ⅰ)⇒(Ⅱ). 设(S, d)是紧的. 给定 $r>0$, $x\in S$, 记 $B(x, r):=\{y\in S: d(x, y)<r\}$, 则对每个 r, 所有邻域$\{B(x, r): x\in S\}$的集合是一个开覆盖, 且一定有一个有限子覆盖. 因此, (S, d)是完全有界的. 现在设$\{x_n\}$是 S 中的任意柯西序列, 则对每一个正整数 m, 存在 $n(m)$, 使得当 $n>n(m)$时, 有 $d(x_n, x_{n(m)})<\frac{1}{m}$. 令 $U_m=\left\{x: d(x, x_{n(m)})>\frac{1}{m}\right\}$, 则 U_m 是一个开集.(如果 $y\in U_m$, $r:=d(x_{n(m)}, y)-\frac{1}{m}$, 则 $r>0$, $B(y, r)\subset U_m$.)由 $n(m)$的定义知, 当 $n>n(m)$时, $x_n\notin U_m$. 因而当 $k>\max\{n(m): m<s\}$时, 有 $x_k\notin\cup\{U_m: 1\leqslant m<s\}$. 因为 U_m 没有有限子覆盖, 所以它们不能构成 S 的一个开覆盖. 于是存在 x, 对所有的 m 有 $x\notin U_m$. 因此, 对所有的 m, 有 $d(x, x_{n(m)})\leqslant\frac{1}{m}$. 则由三角不等式知, 当 $n>n(m)$时, 有 $d(x, x_n)<\frac{2}{m}$, 所以$\lim\limits_{n\to\infty}d(x, x_n)=0$, 并且序列$\{x_n\}$收敛到 x. 故(S, d)是完备的, 并且是完全有界的.

(Ⅱ)⇒(Ⅲ). 对每一个 $n=1, 2, \cdots$, 设 F_n 是 S 的有限子集, 使得对每一个 $x\in S$, $d(x, y)<\frac{1}{n}$ 对某些 $y\in F_n$ 成立. 设 A 是 S 的任意有限子集.(如果 S 是有限的, 则由普通的逻辑知(Ⅲ)显然成立.)因为对于 $y\in F_1$ 有有限个邻域 $B(y, 1)$覆盖 S, 所以一定存在某个 $x_1\in F_1$, 使得$A\cap B(x_1, 1)$是无限的. 由归纳法, 对所有的 n, 取 $x_n\in F_n$, 使得对所有正整数 n, $A\cap\left\{B\left(x_m, \frac{1}{m}\right): m=1, 2, \cdots, n\right\}$ 是无限的. 这就意味着当 $m<n$ 时, $d(x_m, x_n)<\frac{1}{m}+\frac{1}{n}<\frac{2}{m}$ (存在 $y\in B\left(x_m, \frac{1}{m}\right)\cap B\left(x_n, \frac{1}{n}\right)$, 且 $d(x_m, x_n)<d(x_m, y)+d(x_n, y)$). 因而$\{x_n\}$是一个柯西序列. 因为$(S, d)$是完备的, 所以这个柯西序列收敛于 S 中的点 x, 并且对所有的 n, 有 $d(x_n, x)<\frac{2}{n}$, 故 $B\left(x, \frac{3}{n}\right)\supset B\left(x_n, \frac{1}{n}\right)$, 且前者包含了 A 的一个无限子集. 因为当 $n\to\infty$ 时, $\frac{3}{n}\to 0$, 所以 x 是 A 的一个极限点.

(Ⅲ)⇒(Ⅳ). 如果$\{x_n\}$是一个有无限值域的序列, 令 x 是这个值域中的极限点. 那么存在 $n(1)<n(2)<n(3)<\cdots$, 使得对所有的 k, $d(x_{n(k)}, x)<\frac{1}{k}$, 因此, 当 $k\to\infty$ 时, $x_{n(k)}$收敛于 x. 如果$\{x_n\}$有有限值域, 那么存在某个 x, 使得对无限多的 n, 有 $x_n=x$. 因此, 存在一个子序列 $x_{n(k)}$, 对所有的 k, 有 $x_{n(k)}=x$, 故 $x_{n(k)}\to x$.

(Ⅳ)⇒(Ⅰ). 设$\mathcal{U}$是 S 的一个开覆盖, 对任意的 $x\in S$, 令

$$f(x):=\sup\{r: B(x,r)\subset U, U\in\mathcal{U}\},$$

则对每个 $x\in S$, 有 $f(x)>0$. 一个更强的结果成立.

2.3.2 引理　$\operatorname{Inf}\{f(x): x\in S\}>0$.

证明　如果命题不成立, 则在 S 中存在序列$\{x_n\}$, 使得对 $n=1, 2, \cdots$, 有 $f(x_n)<\frac{1}{n}$. $x_{n(k)}$是一个收敛到 S 内某一点 x 的子序列. 则对于某个 $U\in\mathcal{U}$和 $r>0$, $B(x, r)\subset U$. 于是对充分大的 k, 使

得 $d(x_{n(k)}, x) < \frac{r}{2}$，有 $f(x_{n(k)}) > \frac{r}{2}$. 对于足够大的 k 这是一个矛盾．□

现在继续证明(Ⅳ)⇒(Ⅰ)，令 $c := \min(2, \inf\{f(x): x \in S\}) > 0$，选择任意的 $x_1 \in S$，利用递归的方法，给定 $x_1, \cdots, x_n$ 如果可能的话选择 x_{n+1}，使得 $d(x_{n+1}, x_j) > \frac{c}{2}$，对一切 $j = 1, \cdots, n$ 都成立．如果这对所有的 n 都是可能的，我们将得到一个序列 $\{x_n\}$，满足当 $m \neq n$ 时，$d(x_m, x_n) > \frac{c}{2}$. 这样的序列没有柯西子序列，因而没有收敛的子序列．于是存在一个有限的 n，使得 $S = \bigcup_{j \leq n} B(x_j, c/2)$. 由 f 和 c 的定义知，对于每一个 $j = 1, \cdots, n$，存在一个 $U_j \in \mathcal{U}$，使得 $B(x_j, c/2) \subset U_j$. 那么这些 U_j 的并集就是 S，且$\mathcal{U}$有一个有限子覆盖．从而完成了定理 2.3.1 的证明．□

对任意度量空间(S, d)及 $A \subset S$，A 的直径(diameter)定义为 $\operatorname{diam} A := \sup\{d(x, y): x \in A, y \in A\}$. 称 A 是有界的，当且仅当 A 的直径是有限的.

例：设 S 是任意的无限集，对 S 中的 $x \neq y$，令 $d(x, y) = 1$，$d(x, x) = 0$. 则 S 是完备的且是有界的，但不是完全有界．因此，欧几里得空间中紧集的一些特性(比如闭有界性)不能推广到一般的完备度量空间上.

完全有界的度量空间可以根据下述定量标准来衡量完全有界：设(S, d)是完全有界的度量空间，给定 $\varepsilon > 0$，令 $N(\varepsilon, S)$是对某些集合 $A_i (i = 1, 2, \cdots, n)$，$\operatorname{diam}(A_i) \leq 2\varepsilon$，使得$S = \bigcup_{1 \leq i \leq n} A_i$ 的最小 n. 令 $D(\varepsilon, S)$是点 $x_i (i = 1, 2, \cdots, m)$，使得当 $i \neq j$ 时，$d(x_i, x_j) > \varepsilon$ 的最大数 m.

习题

1. 证明：对于任意的度量空间(S, d)及 $\varepsilon > 0$，$N(\varepsilon, S) \leq D(\varepsilon, S) \leq N(\varepsilon/2, S)$.
2. 设(S, d)是具有通常度量的单位区间$[0, 1]$，对于所有的 $\varepsilon > 0$，估计 $N(\varepsilon, S)$和 $D(\varepsilon, S)$．[提示：利用取整函数$[x]$:= 不小于 x 的最小整数．]
3. 如果 S 是$\mathbb{R}^2$ 上具有通常度量的单位正方形$[0, 1] \times [0, 1]$，证明：对于某个常数 K，$N(\varepsilon, S) \leq K/\varepsilon^2$，其中 $0 < \varepsilon < 1$.
4. 给出开单位正方形$(0, 1) \times (0, 1)$的一个开覆盖，它没有有限的子覆盖．
5. 证明：可分度量空间的任意开覆盖有可数个子覆盖．[提示：利用命题 2.1.4.]
6. 证明：一个度量空间是紧的，当且仅当每一个可数滤子基都包含在一个收敛的滤子基中．
7. 由区间$(j/n, (j+2)/n)(j = -1, 0, 1, \cdots, n-1)$覆盖区间$[0, 1]$，计算引理 2.3.2 中的下确界．
8. 设(S, d)是一个非紧的度量空间，于是有无限多的集合 A 没有极限点．证明：A 上的相对拓扑是离散的.
9. 证明：一个具有离散相对拓扑的集合可以有一个极限点．
10. 拓扑空间中的点 x 称为*孤立点*(isolated point)，当且仅当$\{x\}$是开的. 一个紧的拓扑空间称为是*完满的*(perfect)，当且仅当它没有孤立点．证明：
    (a) 任意紧的度量空间是一个可数集和一个完满集的并．[提示：利用习题 5，考虑具有可数个开邻域的点集．]
    (b) 设(K, d)是完满的，则 K 的每一个非空开子集都是不可数的.
11. 设$\{x_i, i \in I\}$是一个网格，其中 I 是一个有向集．对于 $J \subset I$，如果 J 是 I 的*共尾*(cofinal)，即对某个 $j \in J$，对所有的 $i \in I$，有 $i \leq j$，则$\{x_i, i \in J\}$称为$\{x_i, i \in I\}$的*严格子网格*(strict subnet).
    (a) 证明：这就意味着 J 关于 I 的序是一个有向集．
    (b) 证明：在具有通常拓扑的$[0, 1]$上，存在一个网格，它没有收敛的严格子网格(与定理 2.2.5 和定

理 2.3.1 有明显差别). [提示: 设 W 是 $[0, 1]$ 的良序集, I 是使得 $\{t: tWy\}$ 是可数的所有 $y \in [0, 1]$ 的集合. 证明 I 是不可数的且通过 W 是良序集. 令对所有的 $y \in I$, $x_y := y$. 证明 $\{x_y: y \in I\}$ 没有收敛的严格子网格.]

用收敛子网格可以描述紧性, (例如 Kelley, 1955, 定理 5.2), 但这只是对不严格子网格来说, 参见 Kelly(1955, p.70 和问题 2.E).

2.4 函数空间的一些度量

首先介绍三个相对简单的事实.

2.4.1 命题 对于任意的度量空间 (S, d), 如果 $\{x_n\}$ 是一个柯西序列, 则它是有界的(即它的值域是有界的). 如果它有一个收敛的子序列 $x_{n(k)} \to x$, 则 $x_n \to x$. 完备度量空间上的任意闭子集都是完备的.

48

证明 如果对 $m > n$, $d(x_m, x_n) < 1$, 则对所有的 m, $d(x_m, x_n) < 1 + \max\{d(x_j, x_n): j < n\} < \infty$, 于是序列是有界的.

如果 $x_{n(k)} \to x$, 则对给定的 $\varepsilon > 0$, 取 m 使得当 $n > m$ 时, $d(x_n, x_m) < \varepsilon/3$, 取 k 使得 $n(k) > m$, 且 $d(x_{n(k)}, x) < \varepsilon/3$, 则

$$d(x_n, x) < d(x_n, x_m) + d(x_m, x_{n(k)}) + d(x_{n(k)}, x) < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon,$$

因此 $x_n \to x$. 由定理 2.1.3(b) 和 (d) 知, 一个完备度量空间的闭子集是完备的. □

一个非完备度量空间 X 的一个闭子集 F 不一定是完备的, 例如 $F = X$. 下面是完备性的一个经典情况.

2.4.2 命题 $\mathbb{R}$ 关于它的通常度量是完备的.

证明 设 $\{x_n\}$ 是柯西列, 由命题 2.4.1 知, 它是有界的, 因此包含于一个有限区间 $[-M, M]$ 中. 由定理 2.2.1 知, 这个区间是完备的, 于是由定理 2.3.1 知, 它有一个收敛子序列, 故由命题 2.4.1 知, $\{x_n\}$ 是收敛的. □

设 (S, d) 和 (T, e) 是两个任意的度量空间, 容易看出从 S 映射到 T 的函数 f 是连续的, 当且仅当对于所有的 $x \in S$ 和 $\varepsilon > 0$, 存在一个 $\delta > 0$, 使得当 $d(x, y) < \delta$ 时, $e(f(x), f(y)) < \varepsilon$. 如果这对一个固定的 x 成立, 则称函数 f 在 x 点是连续的. 如果对于任意的 $\varepsilon > 0$, 存在 $\delta > 0$, 使得对 S 中任意的 x, y, 当 $d(x, y) < \delta$ 时, $e(f(x), f(y)) < \varepsilon$, 则称 f 从 (S, d) 到 (T, e) 是一致连续的 (uniformly continuous).

例如, 从 $\mathbb{R}$ 到自身的函数 $f(x) = x^2$ 是连续的, 但不是一致连续的. (因为对于一个给定的 ε, 随着 $|x|$ 的增大, δ 一定是变小的.)

在研究可数的笛卡儿积之前, 先讨论有界度量的情形, 我们可以按如下方式来使度量有界, 这里 $[0, \infty) := \{x \in \mathbb{R}: x \geqslant 0\}$.

2.4.3 命题 设 f 是从 $[0, \infty)$ 映射到自身的连续函数, 使得

(1) f 是非递减的, 即当 $x \leqslant y$ 时, $f(x) \leqslant f(y)$.

(2) f 满足次可加性, 即对所有的 $x \geqslant 0$, $y \geqslant 0$, $f(x+y) \leqslant f(x) + f(y)$.

49

(3) $f(x) = 0$ 当且仅当 $x = 0$.

则对于任意的度量空间 (S, d), $f \circ d$ 是一个度量, 且对于从 S 到自身的恒等函数 $g(s) \equiv s$, 从 (S, d) 到 $(S, f \circ d)$, 以及从 $(S, f \circ d)$ 到 (S, d) 是一致连续的.

证明 显然 $0 \leqslant f(d(x, y)) = f(d(y, x))$, 它等于 0 当且仅当对 S 中所有的 x, y, $d(x, y) = 0$.

因为三角不等式,

$$f(d(x,z)) \leqslant f(d(x,y)+d(y,z)) \leqslant f(d(x,y))+f(d(y,z)),$$

所以$f \circ d$是一个度量. 因为对所有的$t>0$, 有$f(t)>0$, 且f是连续非递减的函数, 所以对每一个$\varepsilon>0$, 存在$\delta>0$, 使得当$t<\delta$时, $f(t)<\varepsilon$, 并且如果$f(t)<\lambda:=f(\varepsilon)$, 则$t<\varepsilon$. 因此在两个趋势上都有一致连续性. □

假定对$x>0$, $f''(x)<0$, 则f'是递减的, 于是对任意的x, $y>0$,

$$f(x+y)-f(x)=\int_0^y f'(x+t)\,\mathrm{d}t<\int_0^y f'(t)\,\mathrm{d}t=f(y)-f(0).$$

因而如果$f(0)=0$, 则f是次可加的. 于是存在有界函数f, 满足命题 2.4.3 的条件. 例如$f(x):=x/(1+x)$或者$f(x):=\arctan x$.

2.4.4 命题 对于度量空间的任意序列$(S_n, d_n)(n=1, 2, \cdots)$. 积$S:=\prod_n S_n$关于积拓扑是可度量的, 其中的度量$d(\{x_n\}, \{y_n\}):=\sum_n f(d_n(x_n, y_n))/2^n$, 这里$f(t):=t/(1+t)$, $t>0$.

证明 首先, $f(x)=1-\dfrac{1}{1+x}$, 于是f是非递减的, 且$f''(x)=-\dfrac{2}{(1+x)^3}$, 所以函数$f$满足命题 2.4.3 的三个条件, 因此对每一个$n$, $f \circ d_n$是S_n上的一个度量. 为证明d是度量, 首先令$e_n(x, y):=f(d_n(x_n, y_n))/2^n$. 于是对每一个$n$, e_n是S上的伪度量, 因为$f<1$, 所以对所有的x, y, $d(x, y)=\sum_n e_n(x, y)<1$. 显然, d是非负的、对称的. 对S中任意的x, y和z,

$$d(x,z)=\sum_n e_n(x,z) \leqslant \sum_n e_n(x,y)+e_n(y,z)=d(x,y)+d(y,z).$$

(非负项的重新求和, 参见附录 D.)因此d是伪度量. 如果$x \neq y$, 则对于某个n, $x_n \neq y_n$, 有$d_n(x_n, y_n)>0$, $f(d_n(x_n, y_n))>0$和$d(x, y)>0$. 因此d是一个度量. 对于任意的$x=\{x_n\} \in S$, 积拓扑在点x处有一个由所有集合$N(x, \delta, m):=\{y$: 对所有的$j=1, \cdots, m$, $d_j(x_j, y_j)<\delta\}$组成的邻域基, 其中$\delta>0$, $m=1, 2, \cdots$. 给定$\varepsilon>0$, 当n充分大时, $2^{-n}<\varepsilon/2$. 又因为$(\sum_{1 \leqslant j \leqslant n} 2^{-j})\varepsilon/2<\varepsilon/2$,及对所有的$x$,$f(x)<x$和$\sum_{j>n} 2^{-j}<\varepsilon/2$, 所以对每一个$n$, $N(x, \varepsilon/2, n) \subset B(x, \varepsilon):=\{y: d(x, y)<\varepsilon\}$. 50

反之, 假如给定$0<\delta<1$和n. 由$f(1)=1/2$, $f(x)<1/2$得, $x<1$, 则$x=(1+x)f(x)<2f(x)$. 令$\gamma:=2^{-n-1}\delta$. 如果$d(x, y)<\gamma$, 则对$j=1, \cdots, n$, $f(d_j(x_j, y_j))<2^j\gamma<1/2$. 于是$d_j(x_j, y_j)<2^{j+1}\gamma \leqslant \delta$. 因此有$B(x, \gamma) \subset N(x, \delta, n)$. 因此, x的邻域关于d和积拓扑是相同的, 故d度量了这个拓扑. □

不可数个度量空间(每个超过一个点)的积是不可度量的. 例如, 考虑$\{0, 1\}$关于不可数指标集I的同步积, 换句话说, I的子集的所有指标的集合. 令I的有限子集F通过包含关系是有向的, 则对所有的有限集F, 网格1_F关于积拓扑收敛到 1, 但是没有有限集的示性函数序列$1_{F(n)}$能收敛到 1, 因为$F(n)$的并集是可数的, 而不是整个的I.

因此, 为了得到不可数集合上的实函数的可度量空间, 我们需要限制函数空间或者考虑一个除积拓扑外的拓扑. 这里是一个函数空间: 对于任意紧拓扑空间K, 令$C(K)$是K上所有连续实值函数的空间. 对于$f \in C(K)$, 有$\sup|f|:=\sup\{|f(x)|: x \in K\}<\infty$, 由定理 2.2.3, $f[K]$在$\mathbb{R}$上是紧的. 容易看出, $d_{\sup}(f, g):=\sup|f-g|$是$C(K)$上的一个度量.

设(X, d)是度量空间, 从一个拓扑空间S映射到X的连续函数的集族$\mathcal{F}$称为在$x \in S$处是等度

连续的(equicontinuous)，当且仅当对每个 $\varepsilon>0$，存在 x 的一个邻域 U，使得对于所有的 $y\in U$ 和 $f\in\mathcal{F}$，有 $d(f(x), f(y))<\varepsilon$(这里 U 不依赖于 f). $\mathcal{F}$称为是等度连续的，当且仅当它在每一个 $x\in S$ 处都是等度连续的. 设(S, e)是一个度量空间，如果对每一个 $\varepsilon>0$，存在 $\delta>0$，使得对 S 中所有的 x, y 和$\mathcal{F}$中所有的 f，当 $e(x, y)<\delta$ 时，$d(f(x), f(y))<\varepsilon$，则称 F 是一致等度连续的(uniformly equicontinuous). 根据这个概念，下面将给出一个著名结论的推广(下面的推论 2.4.6).

2.4.5 定理 *如果(K, d)是一个紧度量空间，(Y, e)是一个度量空间，则从 K 映射到 Y 的任意等度连续函数族是一致等度连续的.*

证明 如果不是，则存在 $\varepsilon>0$，$x_n\in K$，$u_n\in K$ 和 $f_n\in\mathcal{F}$，使得对所有的 n，$d(u_n, x_n)<1/n$ 且 $e(f_n(u_n), f_n(x_n))>\varepsilon$. 那么根据 K 中的任意序列有一个收敛的子序列(定理 2.3.1)，我们不妨假定对某个 $x\in K$，有 $x_n\to x$，于是 $u_n\to x$. 由在 x 点的等度连续性，当 n 充分大时，有 $e(f_n(u_n), f_n(x))<\varepsilon/2$，$e(f_n(x_n), f_n(x))<\varepsilon/2$. 因此 $e(f_n(u_n), f_n(x_n))<\varepsilon$，矛盾. □

51

2.4.6 推论 *从一个紧度量空间到任一度量空间上的连续函数都是一致连续的.*

从集合 X 到$\mathbb{R}$上的函数族$\mathcal{F}$称为是一致有界的(uniformly bounded)，当且仅当 $\sup\{|f(x)|: f\in\mathcal{F}, x\in X\}<\infty$. 对任意有界实函数族，正如 $C(K)$对 K 是紧的，设 $d_{\sup}(f, g):=\sup|f-g|$，则 $d_{\sup}$是一个度量.

区间[0, 1]上的函数序列 $f_n(t):=t^n$ 由连续函数组成，并且这个序列是一致有界的：$\sup_n\sup_t|f_n(t)|=1$. 但$\{f_n\}$在 1 处不是等度连续的，因此对 $d_{\sup}$不是完全有界的. 于是有下面的经典特征.

2.4.7 定理(阿尔泽拉–阿斯科利，Arzelà-Ascoli) *设(K, e)是一个紧度量空间，$\mathcal{F}\subset C(K)$，则$\mathcal{F}$关于 $d_{\sup}$是完全有界的，当且仅当它是一致有界的并且是等度连续的，从而是一致等度连续的.*

证明 如果$\mathcal{F}$是完全有界的且 $\varepsilon>0$，取 $f_1, \cdots, f_n\in\mathcal{F}$，使得对所有的 $f\in\mathcal{F}$，对某个 j，有 $\sup|f-f_j|<\varepsilon/3$. 由推论 2.4.6 知，每个 f_j 都是一致连续的. 因此，有限集合$\{f_1, \cdots, f_n\}$是一致等度连续的. 取 $\delta>0$，使得当 $e(x, y)<\delta$ 时，$|f_j(x)-f_j(y)|<\varepsilon/3$ 对所有的 $j=1, \cdots, n$ 及 x, $y\in K$ 成立. 那么对所有的 $f\in\mathcal{F}$，$|f(x)-f(y)|<\varepsilon$，所以$\mathcal{F}$是一致等度连续的. 在任意的度量空间中，一个完全有界的集合是有界的，对 $d_{\sup}$这就意味着是一致有界的.

反之，设$\mathcal{F}$是一致有界和等度连续的，因此，由定理 2.4.5 知，它也是一致等度连续的. 对所有的 $f\in\mathcal{F}$和 $x\in K$，设 $|f(x)|\leqslant M<\infty$. 则由定理 2.2.1 知，$[-M, M]$是紧的. 设$\mathcal{G}$在$\mathcal{F}$在$\mathbb{R}^k$ 的积拓扑中的闭包，则由定理 2.2.8 和定理 2.2.2 知，$\mathcal{G}$是紧的. 对任意的 $\varepsilon>0$ 和 x, $y\in K$，$\{f\in\mathbb{R}^k: |f(x)-f(y)|\leqslant\varepsilon\}$是闭的. 因此，当 $e(x, y)<\delta$ 时，对所有的 $f\in\mathcal{F}$，有 $|f(x)-f(y)|\leqslant\varepsilon$，对所有 $f\in\mathcal{G}$结论同样成立. 故$\mathcal{G}$是一致等度连续的.

设$\mathcal{U}$是$\mathcal{G}$中的任意超滤子，则由定理 2.2.5 知，$\mathcal{U}$(关于积拓扑)收敛于某个 $g\in\mathcal{G}$. 对于给定的 $\varepsilon>0$，取 $\delta>0$，使得当 $e(x, y)<\delta$ 时，$|f(x)-f(y)|\leqslant\varepsilon/4<\varepsilon/3$ 对所有的 $f\in\mathcal{G}$成立. 取一有限集 $S\subset K$，使得对任意的 $y\in K$，对某个 $x\in S$，有 $e(x, y)<\delta$. 令

$$U:=\{f: 对所有的\ x\in S, |f(x)-g(x)|<\varepsilon/3\}.$$

则 U 是$\mathbb{R}^k$ 中的开集，因此 $U\in\mathcal{U}$. 如果 $f\in U$，则 $|f(y)-g(y)|<\varepsilon$ 对所有的 $y\in K$ 成立，故 $d_{\sup}(f, g)\leqslant\varepsilon$. 因此关于 $d_{\sup}$有$\mathcal{U}\to g$. 由定理 2.2.5 知，$\mathcal{G}$关于 $d_{\sup}$是紧的，因此，由定理 2.3.1 知，$\mathcal{F}$是完全有界的. □

52

对于任意的拓扑空间$(S, \mathcal{T})$，设 $C_b(S):=C_b(S, \mathcal{T})$是 S 上所有有界实值连续函数的集合. $d_{\sup}$

是定义在 $C_b(S)$ 上的度量，关于 $d_{\sup}$ 收敛的任意序列 f_n 称为是一致收敛的. 一致收敛保证了有界性和连续性.

2.4.8 定理 对于任意的拓扑空间 $(S, \mathcal{T})$，如果 $f_n \in C_b(S, \mathcal{T})$，且当 $n \to \infty$ 时，一致地有 $f_n \to f$，则 $f \in C_b(S, \mathcal{T})$.

证明 对于任意的 $\varepsilon > 0$，取 n 使得 $d_{\sup}(f_n, f) < \varepsilon/3$，对于任意的 $x \in S$，取 x 的邻域 U，使得对所有的 $y \in U$，$|f_n(x) - f_n(y)| < \varepsilon/3$，则

$$\begin{aligned}|f(x) - f(y)| &\leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\end{aligned}$$

因此 f 是连续的. 因为 $d_{\sup}(0, f) \leqslant d_{\sup}(0, f_n) + d_{\sup}(f_n, f) < \infty$，所以 f 有界. □

2.4.9 定理 对于任意的拓扑空间 $(S, \mathcal{T})$，度量空间 $(C_b(S, \mathcal{T}), d_{\sup})$ 是完备的.

证明 设 $\{f_n\}$ 是柯西序列，则对 S 中的任意 x，$\{f_n(x)\}$ 是 $\mathbb{R}$ 中的柯西序列，于是它收敛到某个实数，记为 $f(x)$. 那么对每一个 m 和 x，当 $m \to \infty$ 时，$|f(x) - f_m(x)| = \lim_{n\to\infty} |f_n(x) - f_m(x)| \leqslant \limsup_{n\to\infty} d_{\sup}(f_n, f_m) \to 0$，因此 $d_{\sup}(f_m, f) \to 0$. 由定理 2.4.8 知，$f \in C_b(S, \mathcal{T})$. □

如果对所有的 n，$c_n \geqslant c_{n+1}$，且当 $n \to \infty$ 时，$c_n \to c$，则记 $c_n \downarrow c$. 如果 f_n 是集合 X 上的实值函数，则 $f_n \downarrow f$ 表示对所有的 $x \in X$，有 $f_n(x) \downarrow f(x)$. 那么有下面的定理.

2.4.10 迪尼定理 设 $(K, \mathcal{T})$ 是一个紧拓扑空间，对所有的 $n \in \mathbb{N}$，f_n 是 K 上的连续实值函数，且 $f_n \downarrow f_0$，则在 K 上一致地有 $f_n \to f_0$.

证明 对每一个 n，$f_n - f_0 \geqslant 0$. 给定 $\varepsilon > 0$，令 $U_n := \{x \in K: (f_n - f_0)(x) < \varepsilon\}$，则 U_n 是开集，且它们的并集是整个 K. 因此，它们有一个有限子覆盖. 因为收敛是单调的，所以有包含关系 $U_n \subset U_{n+1} \subset \cdots$. 因此，某个 U_n 就是整个 K. 则对所有的 $m \geqslant n$，$(f_m - f_0)(x) < \varepsilon$ 对所有的 $x \in K$ 成立. □ [53]

例：在区间 $[0, 1)$ 上函数 $x^n \downarrow 0$ 但不是一致收敛的. $[0, 1)$ 不是紧的，而 $[0, 1]$ 是紧的，$x^n \downarrow I_{\{1\}}$ 不是一致的. 这里的极限函数 f_0 不是连续的. 这也说明了迪尼(Dini)定理中的一些假设是必要的.

集合 X 上的一些实值函数组成的集族 $\mathcal{F}$ 构成向量空间(vector space)，当且仅当对任意的 $f, g \in \mathcal{F}$，$c \in \mathbb{R}$，有 $cf + g \in \mathcal{F}$，其中对所有的 x，$(cf+g)(x) := cf(x) + g(x)$. 此外，如果 $fg \in \mathcal{F}$，这里对所有的 x，$(fg)(x) := f(x)g(x)$，则称 $\mathcal{F}$ 为一个代数(algebra). 如果对在 X 中所有的 $x \neq y$，对某个 $f \in \mathcal{F}$，有 $f(x) \neq f(y)$，则 $\mathcal{F}$ 称为是 X 的一个可分离点(separate point).

2.4.11 斯通–魏尔斯特拉斯定理 设 K 是任意的紧豪斯多夫空间，$\mathcal{F}$ 是包含在 $C(K)$ 中的一个代数，它分离点且包含常数，则 $\mathcal{F}$ 在 $C(K)$ 中关于 $d_{\sup}$ 是稠密的.

定理 2.4.11 有下述推论.

2.4.12 推论(魏尔斯特拉斯) 在任意紧集 $K \subset \mathbb{R}^d (d < \infty)$ 上，有 d 个变量的所有多项式的集合在 $C(K)$ 中关于 $d_{\sup}$ 稠密.

定理 2.4.11 的证明 魏尔斯特拉斯定理的一种特殊情况是很有用的. 对于任意的 $x \in \mathbb{R}$ 和 $k = 1, 2, \cdots$，定义 $\binom{x}{k} := x(x-1)\cdots(x-k+1)/k!$，且 $\binom{x}{0} := 1$. 函数 $t \mapsto (1-t)^{1/2}$ 在 $t = 0$ 处的泰勒展开是一个二项式级数：

$$(1-t)^{1/2} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-t)^n.$$

对任意的 $r < 1$，当 $|t| \leqslant r$ 时，这个级数绝对和一致收敛到该函数(参见附录 B 的例(c)). 因此对

任意的 $\varepsilon>0$，函数

$$(1+\varepsilon-t)^{1/2}=(1+\varepsilon)^{1/2}(1-t/[1+\varepsilon])^{1/2}$$

在[0，1]上有一个一致收敛到它的泰勒级数．令 $\varepsilon\downarrow 0$，我们有

$$\sup_{0\leqslant t\leqslant 1}[(1+\varepsilon-1)^{1/2}-(1-t)^{1/2}]\to 0,$$

因此，函数 $(1-t)^{1/2}$ 在 $0\leqslant t\leqslant 1$ 时可以用多项式一致逼近．令 $t=1-s^2$，函数 $A(s):=|s|$ 在 $-1\leqslant s\leqslant 1$ 时可以用多项式来一致逼近．如果对所有的 x，$|f(x)|\leqslant 1$，令 $(A\circ f)(x):=|f(x)|$．

54 如果 P 是任意多项式且 $f\in\mathcal{F}$，则 $P\circ f\in\mathcal{F}$，其中对所有的 x $(P\circ f)(x):=P(f(x))$．

设 $\overline{\mathcal{F}}$ 是 $\mathcal{F}$ 关于 $d_{\sup}$ 的闭包，由命题 2.4.1 和定理 2.4.9 知，这个闭包等于它的完备化，也包含在 $C(K)$ 中．容易验证，$\overline{\mathcal{F}}$ 也是代数，对任意的 $f\in\overline{\mathcal{F}}$ 和 $M>d_{\sup}(0,f)=\sup|f|$，有 $|f|=MA\circ(f/M)$，因此 $|f|\in\overline{\mathcal{F}}$. 从而对任意的 f，$g\in\overline{\mathcal{F}}$，有

$$\max(f,g)=\frac{1}{2}(f+g)+\frac{1}{2}|f-g|\in\overline{\mathcal{F}},$$

$$\min(f,g)=\frac{1}{2}(f+g)-\frac{1}{2}|f-g|\in\overline{\mathcal{F}},$$

重复上述过程，$\overline{\mathcal{F}}$ 中有限多个函数的最大值和最小值还是在 $\overline{\mathcal{F}}$ 中.

对于 X 中任意的 $x\neq y$，取 $f\in\mathcal{F}$，使得 $f(x)\neq f(y)$，则对任意的实数 c，d，存在 a，$b\in\mathbb{R}$，使得 $(af+b)(x)=c$，$(af+b)(y)=d$，即 $a:=(c-d)/(f(x)-f(y))$，$b:=c-af(x)$，注意到 $af+b\in\mathcal{F}$，现在取任意的 $h\in C(K)$，固定 $x\in K$. 对任意的 $y\in K$，取 $h_y\in\mathcal{F}$，使得 $h_y(x)=h(x)$ 和 $h_y(y)=h(y)$. 给定 $\varepsilon>0$，存在 y 的一个开邻域 U_y，使得对所有的 $v\in U_y$，有 $h_y(v)>h(v)-\varepsilon$. 集合 U_y 构成了 K 的一个开覆盖，且有一个有限子覆盖 $U_{y(j)}$，$j=1,\cdots,n$. 设 $g_x:=\max_{1\leqslant j\leqslant n}h_y(j)$，则 $g_x\in\overline{\mathcal{F}}$，$g_x(x)=h(x)$，且对所有的 $v\in K$，$g_x(v)>h(v)-\varepsilon$.

对每一个 $x\in K$，存在 x 的一个开邻域 V_x，使得对所有的 $u\in V_x$，有 $g_x(u)<h(u)+\varepsilon$. 那么集合 V_x 有一个 K 的有限子覆盖 $V_{x(1)},\cdots,V_{x(m)}$. 设 $g:=\min_{1\leqslant j\leqslant m}g_{x(j)}$，则 $g\in\overline{\mathcal{F}}$，且 $d_{\sup}(g,h)<\varepsilon$. 令 $\varepsilon\downarrow 0$，则 $h\in\mathcal{F}$. □

复数的问题将在附录 B 中给出，对于复数 $z=x+\mathrm{i}y$，定义绝对值为 $|z|=\sqrt{x^2+y^2}$，于是我们得到了一个有界复值函数的度量 $d_{\sup}$. 这里有一个关于复值情况的斯通-魏尔斯特拉斯(Stone-Weierstrass)定理.

2.4.13 推论 *设 $(K,\mathcal{T})$ 是一个紧豪斯多夫空间，设 $\mathcal{A}$ 是连续函数 $K\mapsto\mathbb{C}$ 的代数，它可分离点 K 且包含常数．假定 $\mathcal{A}$ 是自伴的，即当 g，h 是任意实值函数且 $f=g+\mathrm{i}h\in\mathcal{A}$ 时，有 $\bar{f}=g-\mathrm{i}h\in\mathcal{A}$. 则 $\mathcal{A}$ 在 K 上的所有连续复值函数所组成的空间中关于 $d_{\sup}$ 稠密.*

证明 对于任意的 $f=g+\mathrm{i}h\in\mathcal{A}$，其中 $g:=\mathrm{Re}f$，$h:=\mathrm{Im}f$ 是实值的. 我们有 $g=\frac{1}{2}(f+\bar{f})\in\mathcal{A}$，$h=\frac{1}{2\mathrm{i}}(f-\bar{f})\in\mathcal{A}$. 设 $\mathcal{C}$ 是 $\mathcal{A}$ 中所有实值函数的集合，则 $\mathcal{C}$ 是 $\mathbb{R}$ 上的代数．同时也有 $\mathcal{C}=\{\mathrm{Re}f:f\in\mathcal{A}\}$，$\mathcal{C}=\{\mathrm{Im}f:f\in\mathcal{A}\}$. 因此，$\mathcal{C}$ 分离了 K 中的点．它包含了实常数，从而由定理 2.4.11 知，$\mathcal{C}$ 在 K 上的

55 实值连续函数空间中稠密，对任意的 g，$h\in\mathcal{C}$，取 $g+\mathrm{i}h$，就证明了结论． □

例：$\mathcal{A}$ 是自伴的这个假设不能省略．设单位圆 $T^1:=\{z\in\mathbb{C}:|z|=1\}$ 是紧的. 设 $\mathcal{A}$ 是所有多项

式 $z \mapsto \sum_{j=0}^{n} a_j z^j (a_j \in \mathbb{C}, n=0, 1, \cdots)$所组成的集合．那么$\mathcal{A}$是一个代数，满足推论 2.4.13 除自伴之外的所有条件．T^1上的函数$f(z) := \bar{z} = 1/z$不能由一个$\mathcal{A}$上的多项式P_n一致地逼近．对任意这样的P_n，

$$\frac{1}{2\pi}\int_0^{2\pi} |f(e^{i\theta}) - P_n(e^{i\theta})|^2 d\theta \geq 1,$$

因为对$k=0, 1, \cdots$，交叉项$\int_0^{2\pi} e^{-i(k+1)\theta} d\theta = 0$，将$-i$改为$i$，结果也是相同的．

习题

对任意的$M>0$，$\alpha>0$和度量空间(S, d)，设$\mathrm{Lip}(\alpha, M)$是S上所有实值函数的集合，满足对所有的$x, y \in S$，$|f(x)-f(y)| \leq Md(x, y)^{\alpha}$．(若$\alpha=1$，则这样的函数称为利普希茨(Lipschitz)函数．若$0<\alpha<1$，则这样的函数称为满足阶数为α的赫尔德(Hölder)条件．)

1. 设(K, d)是一个紧的度量空间，$u \in K$，证明：对于任意有限的M和$0<\alpha\leq 1$，$\{f \in \mathrm{Lip}(\alpha, M): |f(u)| \leq M\}$关于$d_{\sup}$是紧的．
2. 设$S=[0, 1]$具有通常的度量且$\alpha>1$，证明：$\mathrm{Lip}(\alpha, 1)$仅仅只包含常量函数．
   [提示：对于$0\leq x \leq x+h \leq 1$，
   $$f(x+h)-f(x) = \sum_{1\leq j\leq n} f(x+jh/n) - f(x+(j-1)h/n).$$
   给出右边第j项的绝对值的一个上界，关于j求和，并令$n\to\infty$．]
3. 试求一个从$[0, 1]$映射到自身的连续函数f_n，使得当$n\to\infty$时，逐点地有$f_n\to 0$，但不是一致的．[提示：令$f_n\left(\frac{1}{n}\right)=1$，$f_n(0)\equiv f_n\left(\frac{2}{n}\right)\equiv 0$．](这就说明了迪尼定理中为什么单调收敛$f_n\downarrow f_0$是有用的．)
4. 不应用本节的任何定理证明：$[0, 1]$上的函数$f_n(x) := x^n$在点1处不是等度连续的．
5. 对$i\in I$，设(S_i, d_i)是度量空间，这里I是个有限集，则在笛卡儿积$S=\prod_{i\in I} S_i$上，令$d(x, y) = \sum_i d_i(x_i, y_i)$，
   (a)证明：d是一个度量．
   (b)证明：d度量了d_i拓扑的积．
   (c)证明：(S, d)完备的当且仅当所有的(S_i, d_i)是完备的．

56

6. 证明下面的每一个函数都具命题 2.4.3 中的性质(1)、(2)和(3)：
   (a)$f(x) := x/(1+x)$.
   (b)$f(x) := \tan^{-1}x$.
   (c)$f(x) := \min(x, 1)$，$0\leq x<\infty$.
7. 证明：$[0, 1]$上的函数$x\mapsto \sin(nx)(n=1, 2, \cdots)$在0点不是等度连续的．
8. 从拓扑空间$(S, \mathcal{T})$映射到度量空间(Y, d)的函数f称为是有界的，当且仅当它的值域是有界的．设$C_b(S, Y, d)$是从S映射到Y的所有有界连续函数的集合．对$f, g \in C_b(S, Y, d)$，设$d_{\sup}(f, g) := \sup\{d(f(x), g(x)): x\in S\}$．如果$(Y, d)$是完备的，证明：$C_b(S, Y, d)$关于$d_{\sup}$是完备的．
9. (佩亚诺曲线)．证明：存在从单位区间$[0, 1]$映上到单位正方形$[0, 1]\times[0, 1]$的连续函数f．[提示：设f是分段线性函数序列f_n的极限，$f_1(t)\equiv(0, t)$，
   $$f_2(t) = \begin{cases} (2t, 0), & \text{若 } 0\leq t\leq 1/4 \\ (1/2, 2t-1/2), & \text{若 } 1/4\leq t\leq 3/4 \\ (2-2t, 1), & \text{若 } 3/4\leq t\leq 1. \end{cases}$$
   在第n个阶段，单位正方形被分为$2^n\cdot 2^n=4^n$个相等的小正方形．这里f_n的图像至少要沿着每一个小正方

形的一条边．那么在下一个阶段，在f_n沿着这样一条边的一个区间上，f_{n+1}首先沿着垂直边走到一半，然后沿着平行于初始边的平分线，返回到最后的顶点，就像f_2与f_1的关系那样．证明这种方案可行，使得f_n一致收敛到f.]

10. 证明：对$k=2, 3, \cdots$，存在一个从$[0, 1]$映上到$\mathbb{R}^k$中单位立方体$[0, 1]^k$的连续函数$f^{(k)}$.[提示：在习题9中，对$0\leqslant t\leqslant 1$，令$f^{(2)}(t):=(g(t), h(t)):=f(t)$. 对于任意的$(x, y, z)\in[0, 1]^3$，在$[0, 1]$中存在$t$和$u$，使得$f(u)=(y, z)$，$f(t)=(x, u)$，于是$f^{(3)}(t):=(g(t), g(h(t)), h(h(t)))=(x, y, z)$. 重复这些构造即可．]
11. 证明：存在从$[0, 1]$到$\prod_{n\geqslant 1}[0, 1]_n$的连续映上函数，其中$\prod_{n\geqslant 1}[0, 1]_n$是$[0, 1]$的同步可数积，具有积拓扑．[提示：像在习题10中那样取序列$f^{(k)}$，设
$$F_k(t)_n := \begin{cases} f^{(k)}(t)_n, & \text{对 } n\leqslant k \\ 0, & \text{对 } n>k. \end{cases}$$
证明当$k\to\infty$时，F_k收敛到要证的函数．]
12. 设K是一个紧的豪斯多夫空间，假设对某个k，有k个从K映射到$\mathbb{R}$的连续函数$f_1, \cdots, f_k$，使得$x\mapsto(f_1(x), \cdots, f_k(x))$从$K$映射到$\mathbb{R}^k$是一对一的. 设$\mathcal{F}$是包含$f_1, f_2, \cdots, f_k$和1的最小函数的代数．
(a)证明：$\mathcal{F}$在$C(K)$中关于$d_{\sup}$在是稠密的.
(b)设$K:=S^1:=\{(\cos\theta, \sin\theta): 0\leqslant\theta\leqslant 2\pi\}$是$\mathbb{R}^2$上关于相对拓扑的单位圆．

57

容易证明(a)对于$k=2$是成立的，但是对于$k=1$是不成立的：不存在从S^1映射到$\mathbb{R}$的$1-1$的连续函数．[提示：应用2.2节习题14的(d)及介值定理．对于θ考虑区间$[0, \pi]$和$[\pi, 2\pi]$.]
13. 不用吉洪诺夫定理或滤子证明阿尔泽拉－阿斯科利定理2.4.7的必要性．[提示：应用定理2.4.5，给定$\varepsilon>0$，对$\varepsilon/4$和$\mathcal{F}$，取$\delta>0$，定理2.3.1给出了K中有限的δ-稠密集，且$[-M, M]$有一个有限的$\varepsilon/4$-稠密集，用这些在$\mathcal{F}$中关于$d_{\sup}$构造一个有限的ε-稠密集．]

2.5 度量空间的完备化和完备性

设(S, d)和(T, e)是两个度量空间，一个从S映射到T的函数f称为等距同构(isometry)，当且仅当对所有的$x, y\in S$，有$e(f(x), f(y))=d(x, y)$.

例如，如果$S=T=\mathbb{R}^2$，具有通常的欧几里得距离度量(如2.2节习题15～16)，则等距同构可以通过下面方法获得：取$f(u)=u+v$(变换)，其中v是向量；或者通过围绕任意中心点的旋转；或者通过在任意一条直线上的反射；或者由它们的组合而得到.

下面将证明任意的度量空间S与一个完备集合T的稠密子集是等距的. 一个经典例子是，S是有理数空间$\mathbb{Q}$，$T=\mathbb{R}$. 事实上，这有时用来作为$\mathbb{R}$的定义.

2.5.1 定理 *设(S, d)是任意的度量空间，则存在一个完备的度量空间(T, e)和一个从S映上到T的一个稠密子集的等距同构f.*

注：既然f唯一地保持S上给定的结构(度量)，那么就可以认为S是T的一个子集，此时称T为S的**完备化**(completion).

证明 令$f_x(y):=d(x, y)$，$x, y\in S$. 选一点$u\in S$，令$F(S, d):=\{f_u+g: g\in C_b(S, d)\}$. 在$F(S, d)$上，令$e:=d_{\sup}$. 尽管$F(S, d)$中的函数可能是无界的(如果$S$关于$d$是无界的)，但是它们的差是有界的，且$e$是$F(S, d)$上完全确定的度量．对任意的$x, y, z\in S$，由三角不等式知，$|d(x, z)-d(y, z)|\leqslant d(x, y)$，且等号在$z=x$或$y$时成立．因此，对任意的$z\in S$，$f_z$是连续的，并且对任意的$x, y$，有$f_y-f_x\in C_b(S, d)$和$d_{\sup}(f_x, f_y)=d(x, y)$. 又因为$f_y\in F(S, d)$，且

$F(S, d)$不依赖于u的选择，所以从S映射到$F(S, d)$的函数$x \mapsto f_x$关于d和e是一个等距同构．设T是这个函数在$F(S, d)$中的值域的闭包．因为$(C_b(S, d), d_{sup})$是完备的(定理2.4.9)，因此$F(S, d)$是完备的．从而(T, e)是完备的，它可以作为(S, d)的完备化． □ 58

设(T', e')和f'也满足定理2.5.1的结论．则在f的值域上，$f' \circ f^{-1}$是从T的一个稠密子集映上到T'的一个稠密子集的一个等距同构．因为(T, e)和(T', e')都是完备的，所以这个等距同构自然地拓展到T映上到T'的一个等距同构．因此，(T, e)唯一地被等距同构决定了，从这个意义上我们称它为S的完备化．

如果一个空间一开始就是完备的，比如$\mathbb{R}$，则它的完备化不增加任何新点，这个空间在等距同构的意义下与它的完备化是相等的．

拓扑空间S中的一个集合称为**无处稠密的**(nowhere dense)，当且仅当对每一个非空开集$U \subset S$，存在非空开集$V \subset U$，使得$A \cap V = \varnothing$．回忆一个拓扑空间$(S, \mathcal{T})$称为是可分的，当且仅当S有可数个稠密子集．

例如，在区间$[0, 1]$中，每一个有限集合都是无处稠密的，有限集合的可数并还是可数的．这个并集可能是稠密的，但是它有稠密的补集．这是下面事实的一个例子．

2.5.2 范畴定理 设(S, d)是任意的完备度量空间，$A_1, A_2, \cdots$是S的无处稠密子集序列，则它们的并$\bigcup_{n \geq 1} A_n$有稠密补集．

证明 如果S是空集，则结论显然成立．下面假定S非空，选取$x_1 \in S$，$0 < \varepsilon_1 < 1$，类似地，递归地选择$x_n \in S$，$0 < \varepsilon_n < 1/n$，使得对所有的n，

$$B(x_{n+1}, \varepsilon_{n+1}) \subset B(x_n, \varepsilon_n/2) \setminus A_n.$$

因为A_n是无处稠密的，所以这是可能的．于是对所有的$m \geq n$，有$d(x_m, x_n) < 1/n$，因此$\{x_n\}$是柯西序列．它收敛到某一个x，且对所有的n，有$d(x_n, x) \leq \varepsilon_n/2$，所以$d(x_{n+1}, x) \leq \varepsilon_{n+1}/2 < \varepsilon_{n+1}$且$x \notin A_n$．因为$x_1 \in S$和$\varepsilon_1 > 0$都是任意的，球$B(x_1, \varepsilon_1)$关于拓扑构成了一个基，所以$S \setminus \bigcup_n A_n$是稠密的． □

可数多个无处稠密集的并称为*第一范畴*(first category)集．不是第一范畴的集合称为*第二范畴*(second category)集．(这个术语在数学上与词“范畴”的其他应用没有关系，正如在同调代数中那样)．范畴定理2.5.2是说，每一个完备的度量空间S是第二范畴的．并且如果A是第一范畴的，则$S \setminus A$是第二范畴的．一个度量空间(S, d)称为是**拓扑完备的**(topologically complete)，当且仅当在S上存在某个度量e，与d具有相同的拓扑，使得(S, e)是完备的．因为范畴定理的结论是借助于拓扑给出的，而不依赖于具体的度量，因此，这个定理在拓扑完备空间中也成立．例如，$(-1, 1)$关于通常的度量不是完备的，但是关于度量$e(x, y) := |f(x) - f(y)|$是完备的，这里$f(x) := \tan(\pi x/2)$． 59

由拓扑的定义，任意开集的并或者有限个开集的交是开集．一般来说，可数多个开集的交不一定是开集．这样的集合称为G_δ(源自德语 Gebiet-Durchschnitt)．一个G_δ的补集(即可数多个闭集的并)称为F_σ(源自法语 fermé-somme)．

对于任意的度量空间(S, d)，$A \subset S$和$x \in S$，令

$$d(x, A) := \inf\{d(x, y) : y \in A\}.$$

对于任意的x，$z \in S$及$y \in A$，由$d(x, y) \leq d(x, z) + d(z, y)$，通过在$A$中关于$y$取下确界，给出

2.5.3 $$d(x, A) \leq d(x, z) + d(z, A),$$

并且因为 x 与 z 可以交换，因此，

$$|d(x,A)-d(x,A)|\leqslant d(x,z).$$

这是拓扑完备度量空间的一个性质，例如，它可以应用到$\mathbb{R}$中所有无理数的集合，它初看起来相当不完备.

*2.5.4 **定理**　一个度量空间$(S,\ d)$是拓扑完备的，当且仅当 S 在它的关于 d 的完备化中是一个 G_δ.

证明　根据完备化定理 2.5.1，假设 S 是 T 的一个稠密子集，且$(T,\ d)$是完备的.

“必要性”：假设对 T 中的每一开集 U_n，$S=\bigcap_n U_n$. 对于每个 n 和 $x\in S$，令

$$f_n(x):=\frac{1}{d(x,T\setminus U_n)}.$$

令 $g(t):=\frac{t}{1+t}$，就像在命题 2.4.3 和命题 2.4.4(可数积的度量化)中那样，对 S 中任意的 x，y，令

$$e(x,y):=d(x,y)+\sum_n 2^{-n}g(|f_n(x)-f_n(y)|),$$

则 e 是 S 上的一个度量.

设$\{x_m\}$是 S 中关于 e 的柯西序列，则因为 $d\leqslant e$，$\{x_m\}$也是关于 d 的柯西序列并且关于 d 收敛到某个 $x\in T$. 对每个 n，当 $m\to\infty$ 时，$f_n(x_m)$收敛到某一个 $a_n<\infty$，因此，$d(x_m,\ T\setminus U_n)\to 1/a_n>0$，且对所有的 n，$x\notin T\setminus U_n$，于是 $x\in S$.

对任意集合 F，由(2.5.3)式，函数 $d(\cdot,\ F)$是连续的. 于是在 S 上所有的f_n 是连续的，且关于 d 收敛能推出关于 e 收敛. 因而 d 和 e 可以度量同一个拓扑，且关于 e 有 $x_m\to x$. 因此，S 关于 e 是完备的.

60

“充分性”：设 e 是 S 上的任意度量，且和 d 有相同的拓扑且使得$(S,\ e)$是完备的，对于任意的 $\varepsilon>0$，令

$$U_n(\varepsilon):=\left\{x\in T:\operatorname{diam}(S\cap B_d(x,\varepsilon))<\frac{1}{n}\right\},$$

这里 diam 表示关于 e 的直径，B_d 表示关于 d 的球. 设 $U_n:=\bigcup_{\varepsilon>0}U_n(\varepsilon)$，对于任意的 x，v，如果 $x\in U_n(\varepsilon)$且 $d(x,\ v)<\varepsilon/2$，则 $v\in U_n(\varepsilon/2)$，因此，U_n 在 T 中是开集.

因为 d 和 e 有相同的拓扑，现在对所有的 n，$S\subset U_n$. 如果对所有的 n，有 $x\in U_n$，取 $x_m\in S$，使得 $d(x_m,\ x)\to 0$，则由 U_n 和 $U_n(\varepsilon)$的定义可得，$\{x_m\}$关于 e 也是柯西序列. 因此对某个 $y\in S$，$e(x_m,\ y)\to 0$，故 $x=y\in S$. □

在度量空间$(X,\ d)$中，如果 $x_n\to x$，并且对每个 n，当 $m\to\infty$ 时，有 $x_{nm}\to x_n$，则对某个 $m(n)$，$x_{nm(n)}\to x$，我们可以选择 $m(n)$，使得 $d(x_{nm(n)},\ x_n)<\frac{1}{n}$. 然而这个重复的极限性质却在一些不可度量的拓扑空间中不成立，比如 $2^{\mathbb{R}}$ 关于积拓扑. 例如，在 $2^{\mathbb{R}}$ 中存在有限个 $F(n)$，使得 $1_{F(n)}\to 1_{\mathbb{Q}}$，并且对任意有限 F，存在开集 $U(m)$，使得 $1_{U(m)}\to 1_F$. 但是有下面的命题.

*2.5.5 **命题**　在$\mathbb{R}$中不存在开集序列 $U(1)$，$U(2)$，…，使得在 $2^{\mathbb{R}}$ 中当 $m\to\infty$ 时，有 $1_{U(m)}\to 1_{\mathbb{Q}}$.

证明　假设 $1_{U(m)}\to 1_{\mathbb{Q}}$，设 $X:=\bigcap_{m\geq 1}\bigcup_{n\geq m}U(n)$，则 X 是可数多个稠密开集的交. 因此$\mathbb{R}/X$ 是第一范畴集. 但是，如果 $X=\mathbb{Q}$，则$\mathbb{R}$是第一范畴集，这与范畴定理 2.5.2 矛盾. □

这给出了一个不是拓扑完备的空间的例子.

*2.5.6 **推论** 在$\mathbb{R}$中，$\mathbb{Q}$不是一个G_δ，因而不是拓扑完备的.

下面我们将拓扑完备性以可数笛卡儿积的形式给出，这并不奇怪．例如，由命题2.4.4知，紧度量空间序列的积是可度量的，由吉洪诺夫定理知，它还是紧的，于是由定理2.3.1知(在这种情况下，对任何度量化其拓扑的度量,)它是完备的.

61

2.5.7 定理 设$(S_n, d_n)(n=1, 2, \cdots)$是一个完备度量空间序列，则具有积拓扑的笛卡儿积$\prod_n S_n$关于命题2.4.4的度量$d$是完备的.

证明 在积空间中，序列$\{\{x_{mn}\}_{n\geq 1}\}_{m\geq 1}$是一个柯西序列$\{x_m\}_{m\geq 1}$. 对于任意固定的$n$，当$m, k\to\infty$时，因为$d$是一个非负项的和，所以$f(d_n(x_{mn}, x_{kn}))/2^n\to 0$，因此$d_n(x_{mn}, x_{kn})\to 0$，且$\{x_{in}\}_{i\geq 1}$关于$d_n$是一个柯西序列，于是它在$S_n$中收敛于某一个$x_n$. 因为这些对每一个$n$都成立，因此，积空间中的原始序列关于积拓扑是收敛的，故关于命题2.4.4中的d也是收敛的. □

习题

1. 证明：无处稠密集的闭包是一个无处稠密集．
2. 设(S, d)和(V, e)是两个度量空间，在笛卡儿积$S\times V$上取度量
$$\rho(\langle x,u\rangle,\langle y,v\rangle) = d(x,y) + e(u,v).$$
证明：$S\times V$的完备化和S的完备化与V的完备化的积是等距同构的.
3. 证明：在一个完备的度量空间中，一个含有非空开集的第一类范畴集的补集的交不仅是非空的，而且还是不可数的. [提示：集合$\{x\}$是无处稠密的吗?]
4. 证明：无理数集合$\mathbb{R}/\mathbb{Q}$，关于通常的拓扑($\mathbb{R}$的相对拓扑)是拓扑完备的.
5. 关于通常(相对)拓扑，试定义$\mathbb{R}/\{0, 1\}$上的一个完备度量．
6. 关于通常(相对)拓扑，试定义在$\mathbb{R}/\mathbb{Q}$上的一个完备度量．
7. (a)设(S, d)是一个完备的度量空间，X是S的一个G_δ子集，且关于X上的相对拓扑，Y是X的一个G_δ子集，证明：Y在S中是一个G_δ.
   (b)在一般的拓扑空间上证明同样的问题.
8. 证明：平面$\mathbb{R}^2$不是直线的可数并．(一条直线是一个集合$\{\langle x, y\rangle: ax+by=c\}$，其中$a$，$b$不全为0.)
9. C^1曲线是一个从$\mathbb{R}$映射到$\mathbb{R}^2$的函数$t\mapsto(f(t), g(t))$，这里导数$f'(t)$，$g'(t)$对所有的t都存在并且连续．证明：$\mathbb{R}^2$不是C^1曲线的值域的可数并．[提示：证明在一个有限区间上C^1曲线的值域是无处稠密的.]

62

10. 设(S, d)是任意的非紧度量空间．证明：在S上存在有界连续函数f_n，使得对所有的$x\in S$，$f_n(x)\downarrow 0$，但是f_n并不一致收敛到0. [提示：S既不是紧的，也不是完全有界的.]
11. 证明：一个度量空间(S, d)对度量化其拓扑的每一个度量e是完备的，当且仅当它是紧的. [提示：利用定理2.3.1，对于所有不小于1的正整数$m\neq n$，假定$d(x_m, x_n)\geq\varepsilon>0$. 对任意的正整数$j$，$k\geq 1$，令
$$e_{jk}(x,y) := d(x,x_j) + |j^{-1} - k^{-1}|\varepsilon + d(y,x_k).$$
令$e(x, y):=\min(d(x, y), \inf_{j,k} e_{jk}(x, y))$，证明：对任意的$j$，$k$，$r$和$s$及任意的$x$，$y$，$z\in S$，$e_{js}(x, z)\leq e_{jk}(x, y)+e_{rs}(y, z)$，分$k=r$及$k\neq r$两种情况考虑．]

*2.6 连续函数的扩张

这里的问题是，对于定义在一个拓扑空间S的子集F上的连续实值函数f，何时能把f扩张为S上的连续函数？例如，考虑集合$\mathbb{R}\setminus\{0\}\subset\mathbb{R}$，函数$f(x):=1/x$在$\mathbb{R}\setminus\{0\}$上是连续的，但是不能将它扩张到在0点也连续的连续函数．有界函数$\sin(1/x)$也是一样．正如这些例子所示，除非F是闭

集，否则，一般是不能将一个函数扩张的．如果 F 是闭的，则在度量空间、紧豪斯多夫空间，或者是包含这两个空间的一类空间中，可以证明这种扩张是可能的，我们把这些空间称为正规空间，定义如下．

集合称为是不相交的，当且仅当它们的交集是空集．一个拓扑空间 (S, d) 称为是正规的(normal)，当且仅当对于任意两个不相交的闭子集 E 和 F，存在不相交的开集 U 和 V，使得 $E\subset U$，$F\subset V$. 首先要证明能推出正规性的一些性质．

2.6.1 定理 *每个度量空间 (S, d) 都是正规的．*

证明 对任意的集合 $A\subset S$ 及 $x\in S$，令 $d(x, A):=\inf_{y\in A} d(x, y)$，则由定理 2.5.3 知，$d(\cdot, A)$ 是连续的．对任意不相交闭集 E 和 F，令 $g(x):=d(x, E)/(d(x, E)+d(x, F))$．因为 E 是闭集，所以 $d(x, E)=0$ 当且仅当 $x\in E$，同样可以证明 F 的情况．因为 E 和 F 是不相交的，所以在 g 的定义中，分母不可能为 0，因此 g 是连续的．现在对所有的 x，有 $0\leqslant g(x)\leqslant 1$，且 $g(x)=0\Leftrightarrow x\in E$，$g(x)=1\Leftrightarrow x\in F$．令 $U:=g^{-1}((-\infty, 1/3))$，$V:=g^{-1}((2/3, \infty))$，则 U 和 V 就具有我们想要的性质． [63] □

2.6.2 定理 *每一个紧豪斯多夫空间都是正规的．*

证明 设 E 和 F 是不相交的且是闭的．对每个 $x\in E$ 和 $y\in F$，取开集 U_{xy} 和 V_{yx}，使得 $x\in U_{xy}$，$y\in V_{yx}$ 且 $U_{xy}\cap V_{yx}=\varnothing$．对每一个固定的 y，$\{U_{xy}\}_{x\in E}$ 构成了闭集的一个开覆盖，因而由定理 2.2.2 知，E 是一个紧集．于是对一些有限子集 $E(y)\subset E$，存在有限开覆盖 $\{U_{xy}\}_{x\in E(y)}$．令 $U_y:=\bigcup_{x\in E(y)} U_{xy}$，$V_y:=\bigcap_{x\in E(y)} V_{yx}$．则对每个 y，U_y 和 V_y 都是开的，且 $E\subset U_y$，$y\in V_y$，$U_y\cap V_y=\varnothing$．故 V_y 构成了一个紧集 F 的开覆盖，因此对有限的 $G\subset F$，存在一个开子覆盖 $\{V_y\}_{y\in G}$．令 $U:=\bigcap_{y\in G} U_y$，$V:=\bigcup_{y\in G} V_y$，则 U 和 V 是开集并且不相交，$E\subset U$，$F\subset V$. □

如果原来的连续函数仅有两个值 0 和 1，正如在定理 2.6.1 的证明中对度量空间所做的那样，下面的引理将给出一种扩张，这将有助于更一般的扩张定理的证明．

2.6.3 Urysohn 引理 *对于任意正规拓扑空间 $(S, \mathcal{T})$ 和不相交的闭集合 E，F，在 X 上存在连续实值函数 f，使得对所有的 $x\in E$，有 $f(x)=0$，对所有 $y\in F$，有 $f(y)=1$，且在 X 上 $0\leqslant f\leqslant 1$．*

证明 对每一个二进有理数 $q=m/2^n$，其中 $n=0, 1, \cdots$，$m=0, 1, \cdots, 2^n$，因此 $0\leqslant q\leqslant 1$．首先选择一个唯一的表示，使得 m 是奇数或者 $m=n=0$．对于这样的 q，m 和 n，开集 $U_q:=U_{mn}$ 和闭集 $F_q=F_{mn}$ 可以用如下的方法定义出来：对于 $n=0$，令 $U_0=\varnothing$，$F_0:=E$，$U_1:=X\setminus F$，$F_1:=X$．现在假设对于 $0\leqslant j\leqslant n$，U_{mj} 和 F_{mj} 已经被定义了，使得对于 $r<s$，有 $U_r\subset F_r\subset U_s\subset F_s$．这个结论对于 $n=0$ 也是成立的．令 $q=(2k+1)/2^{n+1}$，则对于 $r=k/2^n$ 及 $s=(k+1)/2^n$，$F_r\subset U_s$．因此，F_r 与闭集合 $X\setminus U_s$ 是不相交的．由正规性，取不相交的开集 U_q 和 V_q，使得 $F_r\subset U_q$，$X\setminus U_s\subset V_q$．令 $F_q:=X\setminus V_q$，则有 $F_r\subset U_q\subset F_q\subset U_s$，因此，所有的 F_q 和 U_q 都被递归地定义了．

令 $f(x):=\inf\{q: x\in F_q\}$，则对所有的 x，有 $0\leqslant f(x)\leqslant 1$，且在 E 上 $f=0$，在 F 上 $f=1$．对任意的 $y\in[0, 1]$，$f(x)>y$ 当且仅当对某些二进有理数 $q>y$，有 $x\in X\setminus F_q$．因此 $\{x: f(x)>y\}$ 是一个开集的并集，因而是开集．又 $f(x)<t$ 当且仅当对于某些二进有理数 $q<t$，有 $x\in E_q$，故 $x\in U_r$，其中的二进有理数 r 满足 $q<r<t$．所以 $\{x: f(x)<t\}$ 也是开集的一个并集，因而是开集．故对任意的开区间 (y, t)，$f^{-1}((y, t))$ 是开的，取并集，则可得到 f 是连续的． [64] □

下面是本节的主要结果：

2.6.4 扩张定理(Tietze-Urysohn) 设$(X, \mathcal{T})$是一个正规拓扑空间，F是X的一个闭子集．那么对于任意的$c \geq 0$以及对下述$\mathbb{R}$的每个子集S，关于通常的拓扑，每个从F映射到S的连续函数f都可以扩张为一个从X映射到S的连续函数g：

(a)$S = [-c, c]$.

(b)$S = (-c, c)$.

(c)$S = \mathbb{R}$.

证明 不妨假定$c = 1$(如果在(a)中$c = 0$，则让$g = 0$)．对于(a)，令$E := \{x \in F: f(x) \leq -1/3\}$，和$H := \{x \in F: f(x) \geq 1/3\}$．因为$E$和$H$是两个不相交的闭集，由引理2.6.3知，在$X$上存在一个连续函数$h$，使得对所有的$x$，有$0 \leq h(x) \leq 1$，且在$E$上$h = 0$，在$H$上$h = 1$．令$g_0 := 0$，$g_1 := (2h-1)/3$，则$g_1$是$X$上的连续函数，且对所有的$x$，有$|g_1(x)| \leq 1/3$，$\sup_{x \in F} |f - g_1|(x) \leq 2/3$．依此类推，可以证明，对$n = 1, 2, \cdots$，存在$g_n \in C_b(X, \mathcal{T})$，使得对每个$n$，有

2.6.5 $$\sup_{x \in F} |f - g_n|(x) \leq 2^n/3^n,$$

2.6.6 $$\sup_{x \in X} |g_{n-1} - g_n|(x) \leq 2^{n-1}/3^n.$$

这两个不等式在$n = 1$时成立．设$g_1, \cdots, g_n$对$j = 1, 2, \cdots, n$使得2.6.5式和2.6.6式成立．利用g_1的选择办法以$(3/2)^n(f - g_n)$代替f，这通过2.6.5式可以做到．故存在一个$f_n \in C_b(X, \mathcal{T})$，使得$\sup_{x \in F} |(3/2)^n(f - g_n) - f_n|(x) \leq 2/3$，并且有$\sup_{x \in X} |f_n(x)| \leq 1/3$．令$g_{n+1} := g_n + (2/3)^n f_n$，于是在$n+1$处2.6.5式和2.6.6式仍然成立.

现在当$n \to \infty$时，g_n在X上一致收敛于一个函数g，使得在F上有$f = g$. 对所有的$x \in X$，有$|g(x)| \leq \sum_{1 \leq n < \infty} 2^{n-1}/3^n \leq 1$. 由于每个$g_n$是连续的，由定理2.4.8知，$g$是连续的，则证明了(a).

现在考虑(b)，还是假设$c = 1$，首先应用(a)．然后令$G := |g|^{-1}(\{1\})$，这是一个与F不相交的闭集．由引理2.6.3，取$h \in C_b(X, \mathcal{T})$，使得在$G$上$h = 0$，在$F$上$h = 1$，且在$X$上$0 \leq h \leq 1$. 令$j := hg$，于是$j$是连续的，在$F$上$j = f$，且对所有的$x$，$|j(x)| < 1$，则证明了(b).

对于(c)，只要对函数$2(\arctan f)/\pi$应用(b)，然后取$\tan(\pi g/2)$就可完成证明. □

习题

65

1. (a)证明：$\mathbb{R}$中的开集U是可数多个不相交的开区间的并集，它们中的一个或者两个可以是无界的(直线或者半直线)．[提示：对每一个$x \in U$，在U中求一个最大的包含x的开区间．]

   (b)设F是$\mathbb{R}$中的闭集，由(a)知，它的补集是不相交开区间(a_n, b_n)的并集．设f是F上的连续实值函数，将f扩张为$[a_n, b_n]$上的线性函数，或者如果a_n或b_n是无限的，将f扩张为半直线上的常数．证明：f在$\mathbb{R}$上是连续的．[提示：注意a_n的子序列可能是收敛的.]

2. 设(S, d)是任意的不完备度量空间，证明：存在不相交的闭集E和F，及点$x_n \in E$，$y_n \in F$，使得$d(x_n, y_n) \to 0$. 引理2.6.3中的函数f不是一致连续的．[提示：证明存在一个不收敛的柯西序列$\{z_n\}$，且当$m \neq n$时，$z_m \neq z_n$. 令$E = \{x_n\}$，$F = \{y_n\}$是$\{z_n\}$的子序列．]

3. 证明：如果$a < b$，不能把从两点集$\{a, b\}$映上到自身的示性函数扩张为从$[a, b]$映上到$\{a, b\}$的一个连续函数．[提示：利用连通性，参见2.2节习题14.]

4. 拓扑空间$(S, \mathcal{T})$称为是完满正规的(perfectly normal)，当且仅当对每个闭集F，在S上存在一个连续实值函数f，使得$F = f^{-1}(\{0\})$.

(a)证明：每个度量空间都是完备正规的.

(b)证明：每个完备正规空间都是正规的. [提示：仿照定理 2.6.1 的证明.]

5. 设$(S, \mathcal{T})$是一个完备正规空间，F 是 S 的一个闭子集. 在这个空间中定理 2.6.4 自由(a)推(b)的另一种证明方法如下：证明存在一个 $h \in C_b(S, \mathcal{T})$，使得对每个从 F 映射到$(-1, 1)$的连续函数f和从 S 映射到$[-1, 1]$的连续函数 g，在 F 上它等于f，hg 是一个从 S 映射到$(-1, 1)$的连续函数，且在 F 上等于f. [提示：在 F 上取 $h=1$，在其他地方，$0 \leqslant h < 1$.]

6. 设$(S, \mathcal{T})$是一个正规空间，F 是 S 的一个闭子集，f是一个从 F 映射到$\mathbb{R}^k$ 的闭子集 Y 的连续函数，关于 Y 上的通常(相对)拓扑，证明：下列条件成立时，f能扩张为从 F 映射到 Y 的一个连续函数.

(a) Y 是闭单位球$\{y \in \mathbb{R}^k : \sum_{j=1}^{k} y_j^2 \leqslant 1$. [提示：首先利用定理 2.6.4(a)把每个坐标函数f_j 扩张为 g_j 以得到函数 g，若 $|g(x)| \leqslant 1$，则令 $h(x) := g(x)$，否则 $h(x) := g(x)/|g(x)|$.]

(b) Y 是开单位球$\{y \in \mathbb{R}^k : \sum_{j=1}^{k} y_j^2 < 1\}$. [提示：利用(a)并参看定理 2.6.4(b)的证明.]

66

7. 设(S, d)是一个度量空间，F 是 S 中的一个闭子集，f是 F 上的连续有界非负函数. 对 $x \in S$，令

$$g(x) := \begin{cases} \sup\limits_{t \in F} f(t)/(1 + d(t,x)^2)^{1/d(x,F)}, & \text{若 } x \notin F \\ f(x), & \text{若 } x \in F. \end{cases}$$

证明：g 是 S 上的有界连续函数. [提示：证明 g 既是上半连续又是下半连续的. 参见 2.2 节的习题 18. $\limsup\limits_{y \to x} g(y) \leqslant g(x) \leqslant \liminf\limits_{y \to x} g(y)$. (a)对 $x \notin F$ (b)对 $x \in F$，$y \notin F$. 在情况(b)中，当 $y = y_n \to x$ 时，考虑 $t_n \in F$，使得当 $n \to \infty$ 时，$d(y_n, t_n)/d(y_n, F) \to 1$.

8. 一个紧的豪斯多夫空间是正规的(定理 2.6.2)，但是这样的空间的子集不一定是正规的：令$(A, \leqslant)$是一个可数的良序集，使得仅有一个 $a \in A$，对无限多个 $x \in A$，满足 $x < a$(即 a 是 A 的最大元素). 设$(B, \leqslant)$是一个不可数的良序集，使得只存在一个 $b \in B$，对不可数个 $y \in B$，满足 $x < b$(即 b 是 B 的最大元素).

(a)证明：在 2.2 节的习题 9 中关于它们的区间拓扑，A 和 B 是紧的，则 $A \times B$ 关于积拓扑是紧的，也是正规的(它也被称为 Tychonoff plank).

(b)证明：$A \times B$ 删掉上顶点$\langle a, b\rangle$后不是正规的. [提示：证明闭集 $E := \{a\} \times (B \setminus \{b\})$和 $F := (A \setminus \{a\}) \times \{b\}$是不可分的. 如果 U 是包含 F 的开集，证明对于每一个 $x < a$，对不在某个可数集 B_x 中的所有 y，$(x, y) \in U$. 设 C 是这些 B_x 的并. 对所有的 $x < a$，和任意的 $y \notin C$，$(x, y) \in U$. 但是包含 E 的开集 V 一定含有(x, y)这样一些点.]

9. 设 C_r 是圆周 $x^2 + y^2 = r^2$，D 是单位圆盘 $x^2 + y^2 \leqslant 1$，证明：从 C_1 映上到自身的示性函数不能扩张为从 D 映上到 C_1 的连续函数. [提示：对于任意的点 $p \in D$，取极坐标$(r(p), \theta(p))$，于是 $r \circ g \equiv 1$. 令(r, φ)是在 D 中点的坐标，其中 $0 \leqslant \varphi < 2\pi$. 对于 $r > 0$，证明：$\theta(g(r, \varphi))$可以定义为 φ 的一个连续函数(在$[0, 2\pi)$中不必有定义). 令 $h(r) := \lim\limits_{\varphi \to 2\pi} \theta(g(r, \varphi)) - \theta(g(r, 0))$. 证明：$h$ 作为 r 的函数是连续的，且总是 2π 的倍数，但是在 $r=0$ 和 $r=1$ 处有不同的值.]

*2.7 一致性与一致空间

一致空间具有度量空间的一些性质，因此在这些空间之间可以定义一致连续函数. 首先需要下面的概念. 设 A 是笛卡儿积 $X \times Y$ 的子集，B 是 $Y \times Z$ 的子集，于是 $A \circ B$ 是在 $X \times Z$ 中所有形如$\langle x, z\rangle$的点的集合，使得对某个 $y \in Y$，既有$\langle x, y\rangle \in A$，又有$\langle y, z\rangle \in B$. 这个概念是通常的函数复合概念的推广，其中 y 是唯一的.

67

定义 给定一个集合 S，S 上的**一致结构**(uniformity)是指 $S \times S$ 中具有下列性质的滤子$\mathcal{U}$.

(a)每个 $A \in \mathcal{U}$都包含对角线 $D := \{\langle s, s\rangle : s \in S\}$.

(b)对每个 $A\in\mathcal{U}$，有 $A^{-1}:=\{\langle y, x\rangle: \langle x, y\rangle\in A\}\in\mathcal{U}$.

(c)对每个 $A\in\mathcal{U}$，存在一个 $B\in\mathcal{U}$，使得 $B\circ B\subset A$.

$\langle S, \mathcal{U}\rangle$ 称为**一致空间**(uniform space).

一个集合 $A\subset S\times S$ 称为是对称的(symmetric)，当且仅当 $A=A^{-1}$.

回忆伪度量 d，它满足除了"对某些 $x\neq y$，可能有 $d(x, y)=0$"之外度量的所有性质．设 (S, d) 是一个(伪)度量空间，则关于 d 的伪度量一致结构(pseudo metric uniformity)是 $S\times S$ 的所有子集 A 组成的集类$\mathcal{U}$，使得对某个 $\delta>0$，A 包含 $\{\langle x, y\rangle: d(x, y)<\delta\}$，事实上，容易验证这就是一致性.

设 $\langle S, \mathcal{U}\rangle$ 和 $\langle T, \mathcal{V}\rangle$ 是两个一致空间，则从 S 映射到 T 的函数 f 称为关于这些一致结构是一致连续的，当且仅当对每一个 $B\in\mathcal{V}$，$\{\langle x, y\rangle\in S\times S: \langle f(x), f(y)\rangle\in B\}\in\mathcal{U}$. 如果 T 是实直线，则一般认为$\mathcal{U}$是用通常度量 $d(x, y):=|x-y|$ 定义的一致性结构.

对于任意的一致空间 $\langle S, \mathcal{U}\rangle$，由$\mathcal{U}$所定义的 S 上的一致拓扑$\mathcal{T}$是所有 $V\subset S$ 所组成的集类，满足对每个 $x\in V$，存在一个 $U\in\mathcal{U}$，使得 $\{y: \langle x, y\rangle\in U\}\subset V$.

因为一致结构是一个滤子，一致结构的基恰好就是滤子基，就像在定理 2.1.2 中定义的那样．如果 (S, d) 是一个度量空间，则显然用度量一致结构所定义的拓扑就是用度量所定义的一般拓扑．(伪)度量一致结构描述如下.

2.7.1 定理 一个空间 S 上的一致结构$\mathcal{U}$是可伪度量的(这是关于某个伪度量 d 的伪度量一致结构)，当且仅当$\mathcal{U}$有一个可数基.

证明 "充分性"：集合 $\{\langle x, y\rangle: d(x, y)<1/n\}$ $(n=1, 2, \cdots)$ 显然构成了伪度量 d 一致结构的一个可数基.

"必要性"：设一致结构$\mathcal{U}$有一个可数基 $\{U_n\}$，对于一致结构$\mathcal{U}$中的任意 U，应用(c)两次，存在一个 $V\in\mathcal{U}$，使得 $(V\circ V)\circ(V\circ V)\subset U$. 递推地，令 $V_0:=S\times S$，对每一个 $n=1, 2, \cdots$，令 W_n 是满足"有一个集合 $V\in\mathcal{U}$，满足 $(V\circ V)\circ(V\circ V)\subset V_{n-1}$"的 U_n 的交．令 $V_n:=W_n\cap W_n^{-1}\in\mathcal{U}$，则 $\{V_n\}$ 是 U 的一个基，它由对称集合组成，满足对每个 $n\geqslant 1$，有 $V_n\circ V_n\circ V_n\subset V_{n-1}$. 下面的结果将给出定理 2.7.1 的一个证明．

68

2.7.2 引理 对于上述的 V_n，在 $S\times S$ 上有一个伪度量 d，满足对所有 $n=1, 2, \cdots$，有

$$V_{n+1}\subset\{\langle x,y\rangle: d(x,y)<2^{-n}\}\subset V_n.$$

证明 对 $n=0, 1, \cdots$，设

$$r(x,y):=\begin{cases}2^{-n}, & 若\langle x,y\rangle\in V_n\setminus V_{n+1}\\ 0, & 若\langle x,y\rangle\in V_n.\end{cases}$$

因为每个 V_n 是对称的，所以对 S 中所有的 x，y，有 r：$r(x, y)=r(y, x)$.

对于 S 中的每个 x，y，令 $d(x, y)$ 是关于 $n=0, 1, 2, \cdots$ 的所有和 $\sum_{0\leqslant i\leqslant n} r(x_i, x_{i+1})$ 的下确界，且序列 $x_0, \cdots, x_{n+1}$ 在 S 中，满足 $x_0=x$，$x_{n+1}=y$，则 d 是非负的．因为 r 是对称的，所以 d 也是对称的．由 d 的定义，它满足三角不等式，于是 d 是一个伪度量．由于 $d\leqslant r$，显然，$V_{n+1}\subset\{\langle x, y\rangle: d(x, y)<2^{-n}\}$. 下一步如下.

2.7.3 引理 对于任意的 $x_0, \cdots, x_{n+1}$，有 $r(x_0, x_{n+1})\leqslant 2\sum_{0\leqslant i\leqslant n} r(x_i, x_{i+1})$.

证明 关于 n 进行数学归纳法．对于 $n=0$，结论显然成立．当 $0\leqslant j<k\leqslant n+1$ 时，令$L(j, k):=$

$\sum_{j\leqslant i<k} r(x_i, x_{i+1})$，$L:=L(0, n+1)$. 设 k 是所有不超过 n 的满足 $L(0, i)\leqslant L/2$ 的最大的 i. 于是 $L(k+1, n+1)=L-L(0, k+1)\leqslant L/2$ 且 $k<n+1$. 由归纳假设，$r(x_0, x_k)$ 与 $r(x_{k+1}, x_{n+1})$ 每一个至多为 $2(L/2)=L$，并且显然，$r(x_k, x_{k+1})\leqslant L$. 设 m 是满足 $2^{-m}\leqslant L$ 的最小整数，则 $\langle x_0, x_k\rangle\langle x_k, x_{k+1}\rangle$ 和 $\langle x_{k+1}, x_{n+1}\rangle$ 都属于 V_m. 如果 $m=0$，则引理显然成立，下面假设 $m\geqslant 1$. 由 V_j 的选择，我们有 $\langle x_0, x_{n+1}\rangle\in V_{m-1}$，故 $r(x_0, x_{n+1})\leqslant 2^{1-m}\leqslant 2L$. 即证明了引理 2.7.3. □

因而，如果 $d(x, y)\leqslant 2^{-n}$，则 $r(x, y)<2^{1-n}$，事实上，$r(x, y)\leqslant 2^{-n}$，$\langle x, y\rangle\in V_n$. 这就证明了引理 2.7.2，进而证明了定理 2.7.1. □

一致结构自然地也会出现在拓扑群的研究中，一个**群**(group)是由一个集合 G 与一个从 $G\times G$ 映射到 G 的函数 $(x, y)\mapsto xy$ 构成的. 群满足结合律：$(xy)z\equiv x(yz)$，并且有一个单位元 $e\in G$，对所有的 $x\in G$. 满足 $ex=xe=x$，并且对所有的 $x\in G$，有一个逆元 $y\in G$，使得 $xy=yx=e$，记作 $y=x^{-1}$. 一个**拓扑群**(topological group)是由一个群 G 与 G 上拓扑 $\mathcal{T}$ 构成的，使得从 $G\times G$ 映上到 G 的乘法 $\langle x, y\rangle\mapsto xy$ 是连续的，且从 G 映上到自身的逆运算 $x\mapsto x^{-1}$ 是连续的. 对于任意的拓扑群 $(G, \mathcal{T})$，对于 G 中的单位元的邻域 U，所有集合 $\{\langle x, y\rangle: xy^{-1}\in U\}$ 的集类是一个一致结构的基，称为**右一致结构**(right uniformity). 同样地，集合 $\{\langle x, y\rangle: x^{-1}y\in U\}$（其中 U 是单位元的任意邻域）构成了一致结构的基，称为**左一致结构**(left uniformity).

69

一个群称为**阿贝尔群**(Abelian group)，如果对 G 中的每个 x，y，有 $xy=yx$. 显然，对于任意的阿贝尔群，右一致结构与左一致结构是相同的. 群运算将加法运算代替乘法运算，同样以 $-x$ 代替 x^{-1}.

习题

1. 如果 X，Y 与 Z 是一致空间，f 是从 X 映射到 Y 的一致连续函数，g 是从 Y 映射到 Z 的一致连续函数，证明：$g\circ f$ 是从 X 映射到 Z 的一致连续函数.
2. 证明：任意的一致结构有一个由对称集合组成的基.
3. 对于具有通常拓扑的实直线，求两个给定拓扑的不同的一致结构.[提示：给出拓扑的两个度量，使得单位元从一个度量到另一个度量上不是一致连续的.]
4. 一个一致空间 $(S, \mathcal{U})$ 称为是可分的，当且仅当对于 S 中的每个 $x\neq y$，存在 $U\in\mathcal{U}$，使得 $\langle x, y\rangle\notin U$.
   (a)证明：可分一致结构的一致拓扑总是豪斯多夫的.
   (b)证明：如果 S 的点超过一个，那么一个可分一致结构作为一个滤子关于其积拓扑不收敛.
5. 给定一个一致空间 $(S, \mathcal{U})$，S 中的一个网格 $\{x_\alpha\}_{\alpha\in I}$ 称为**柯西网格**(Cauchy net)，当且仅当对于任意的 $V\in\mathcal{U}$，$\exists\gamma\in I$，使得当 $\alpha\geqslant\gamma$，$\beta\geqslant\gamma$ 时，有 $\langle x_\alpha, x_\beta\rangle\in V$. 如果一致空间的每一个柯西网格（关于 $\mathcal{U}$ 的拓扑）收敛到 S 中的某个元素，则称该一致空间为**完备的**. 证明：如果一个度量空间 (S, d) 是完备的，那么它作为一致空间也是完备的.
6. 证明：任意紧一致空间 S（关于一致结构的拓扑是紧的）是完备的. [提示：对于一个柯西网格 $\{x_\alpha\}$，所有 $\{x_\alpha: \alpha\geqslant\gamma\}$ $(\gamma\in I)$ 所组成的集类是一个能扩张为一个滤子和一个超滤子（定理 2.2.4）的滤子基，并且是收敛的（定理 2.2.5）. 证明：网格 x_α 收敛到同一个极限.]
7. 设 $(S, \mathcal{U})$ 和 $(T, \mathcal{V})$ 是紧一致空间，f 是一个从 S 映射到 T 的函数，关于一致结构的拓扑是连续的. 证明：它是一致连续的.

70

8. 参考习题 7，证明：一个紧的豪斯多夫空间关于它的拓扑有唯一的一致结构.
9. 设 G 是**直线的仿射群**(affine group of the line)，即 $\mathbb{R}$ 映上到它的自身的所有变换 $x\mapsto ax+b$ $(a\neq 0, b\in\mathbb{R})$ 所组成的集合. 设 G 在 $\mathbb{R}^2$ 中有形如 $\{\langle a, b\rangle: a\neq 0\}$ 的相对拓扑. 证明：G 是一个拓扑群，对于该拓扑群，左

一致结构和右一致结构是不同的.

10. 一个阿贝尔群 G 上的度量 d 称为是一个不变度量(invariant metric),当且仅当对 G 中所有的 x, y, z, 有 $d(x+z, y+z)=d(x, y)$.

 (a) 证明:一个可度量的阿贝尔拓扑群可以由一个不变度量 d 所度量.

 (b) 如果一个阿贝尔拓扑群 G 可以由一个完备的度量所度量,证明:对于(a)中的不变度量它也是完备的.[提示:如果不是,取 G 关于 d 的完备化,证明它有一个可连续扩张到 G 的阿贝尔群结构,应用定理 2.5.4 和范畴定理 2.5.2.]

*2.8 紧化

本节的问题是,给定一个拓扑空间$(X, \mathcal{T})$,它是否同胚于一个紧豪斯多夫空间的子集?如果 K 是一个紧的豪斯多夫空间,f 是一个从 X 映上到 K 的稠密子集的同胚,则 K 或者(K, f)称为是 X 或者$(X, \mathcal{T})$的紧化(compactification). 对局部紧空间(也就是每个点都有一个紧邻域)存在一个紧化.

2.8.1 定理 *如果$(X, \mathcal{T})$是一个局部紧(但不是紧的)豪斯多夫空间,则它有一个紧化(K, f),这里 f 的值域包含了 K 所有的点而不是一个点.*

证明 设∞是不在 X 中的任意一点,$K:=X\cup\{\infty\}$,令$\mathcal{U}$是 K 上的拓扑,具有由$\mathcal{T}$中的集合所给出的一个基和对 X 所有紧子集 L 的所有集合$\{\infty\}\cup(X\setminus L)$. 函数 f 是从 X 映射到 K 的恒等函数,容易看出,拓扑$\mathcal{U}$是紧的并且 X 上的相对拓扑是$\mathcal{T}$,因此,f 是一个同胚. □

由定理 2.8.1 所给出的$(X, \mathcal{T})$的紧化$(X\cup\{\infty\}, \mathcal{U})$称为是$(X, \mathcal{T})$的一点紧化(one-point compactification). 例如,如果 $X=\mathbb{R}$ 具有通常的拓扑,那么可以验证,$\mathbb{R}$ 的一点紧化同胚于一个圆(留作习题). 如果把这个点记为 ω,则当 $n\to\infty$(通常意义下的整数变大)时,有 $n\to\omega$(在一点紧化的拓扑中),且有 $-n\to\omega$. 实际上,对于$\mathbb{R}$,另一个紧化更加有用:把两个点 $-\infty$ 与 $+\infty$ 添加到$\mathbb{R}$,这里对于 $n=1, 2, \cdots$, $-\infty$ 的每个邻域包含一些集合$[-\infty, -n)$, $+\infty$ 的每个邻域包含一些集合$(n, +\infty]$. 这个结果就是所谓的实数扩张集合$[-\infty, +\infty]$. 71

对于度量空间,人们可能会问是否存在可度量的紧化. 对于$\mathbb{R}$和许多其他的空间,一点紧化都是可度量的. 但是也有一些空间,一点紧化对它们所起的作用非常小,例如,第 5 章中定义的无限维的巴拿赫空间和希尔伯特空间. 下面举一个例子,设 ℓ^1 是所有满足 $\sum_{n=1}^{\infty}|x_n|<\infty$ 的实数序列 $\{x_n\}$ 所组成的集合. 对于两个这样的序列,令 $d(\{x_n\}, \{y_n\}):=\sum_n|x_n-y_n|$,则$(\ell^1, d)$是一个度量空间,并且是可分的(利用当 n 充分大时,$r_n=0$ 的有理数序列 r_n),(ℓ^1, d)不是局部紧的. 例如,在 0 的任意基本邻域 U 中,$U:=\{\{y_n\}:\sum_n|y_n|<\delta\}$,其中 $\delta>0$,令序列 e_m 满足当 $n=m$ 时,$e_{mn}=\delta/2$;当 $n\neq m$ 时,$e_{mn}=0$,则 e_m 没有收敛的子序列. 事实上,上面所定义的拓扑对于 ℓ^1 的一点紧化是紧的,但是它不是一个豪斯多夫空间.

由于紧度量空间是可分的(因为由定理 2.3.1 知,它是完全有界的),且可分度量空间的任意子集还是可分的(2.1 节的习题 5),度量空间的可度量紧化的一个必要条件是可分性. 这个条件实际上也是充分的(例如,ℓ^1 有一个可度量的紧化).

2.8.2 定理 *对于任意的可分度量空间(S, d),存在一个完全有界的度量化. 即存在 S 上的一个度*

量 e，和 d 一样定义了同一个拓扑，使得 (S, e) 是完全有界的，因此，对于 e，S 的完备化是一个紧度量空间和 S 的紧化.

证明 设 $\{x_n\}_{n\geq 1}$ 在 S 中稠密，令 $f(t) := t/(1+t)$，因而 $f \circ d$ 有一个界为 1 的度量，仿照命题 2.4.3 的证明，它与 d 有相同的拓扑. 因此，我们可以假定 $d < 1$. $[0, 1]$ 的同步笛卡儿积 $\prod_{n=1}^{\infty}[0, 1]$ 满足积拓扑，由吉洪诺夫定理 2.2.8 知，它是紧的. 由命题 2.4.4 知，该拓扑的一个度量是

$$\alpha(\{u_n\}, \{v_n\}) := \sum_n |u_n - v_n| / 2^n.$$

因此，由定理 2.3.1 知，这个度量是完全有界的. 通过 $e(x, y) := \alpha(\{d(x, x_n)\}, \{d(y, x_n)\})$ 在 S 上定义一个度量 e，则 (S, e) 是完全有界的. 现在 S 中的一个序列 $y_m \to y$，当且仅当对所有的 n，$\lim_{m\to\infty} d(y_m, x_n) = d(y, x_n)$：充分性是显然的，必要性可通过取靠近 y 的 x_n 来证明. 于是就像在命题 2.4.4 中那样，$y_m \to y$ 当且仅当 $e(y_m, y) \to 0$. 因此，e 度量了 d 拓扑. (S, e) 的完备化还是完全有界的，故由定理 2.3.1 知，它是紧的. □

72

对于一个一般的豪斯多夫空间 $(X, \mathcal{T})$，它可能不是局部紧的，或者不是可度量的，则紧化的存在性可以如下等价地证明：$(X, \mathcal{T})$ 称为是完全正则的(completely regular)，如果对 X 中的每个闭集 F，和不在 F 中的点 p，在 X 上存在一个实值连续函数 f，使得对所有的 $x \in F$，有 $f(x) = 0$，且 $f(p) = 1$. 例如，取一个紧的豪斯多夫空间 K，并且去掉一点 q，容易证明剩下的空间是完全正则的. 因为 $F \cup \{q\}$ 在 K 中是闭的，由定理 2.6.2 和引理 2.6.3 知，在 K 上存在一个连续函数 f，使得在 K 上，$f(p) = 1$，而在 $F \cup \{q\}$ 上，$f \equiv 0$.

2.8.3 定理(吉洪诺夫) 一个拓扑空间 $(X, \mathcal{T})$ 同胚于一个紧豪斯多夫空间的一个子集，当且仅当 $(X, \mathcal{T})$ 是豪斯多夫空间并且是完全正则的.

注： 一个豪斯多夫完全正则的空间称为是一个 $T_{3\frac{1}{2}}$ 空间，或者是吉洪诺夫空间. 因此，定理 2.8.3 是说，一个空间同胚于一个紧豪斯多夫空间的一个子集，当且仅当它是吉洪诺夫空间.

证明 设 X 是一个紧豪斯多夫空间 K 的一个子集，F 是 X 的一个(相对)闭子集，且 $p \in X \setminus F$. 设 H 是 F 在 K 中的闭包，则 H 是 K 的一个闭子集且 $p \notin H$，同样，$\{p\}$ 在 K 中也是闭的. 于是由定理 2.6.3 和定理 2.6.4 知，由 Tietze-Urysohn 定理 2.6.4 知，p 能被一个连续实值函数 f 从 H 中分离出来. 再把 f 限制到 X 上，然后证明 X 是一个吉洪诺夫空间.

反之，设 X 是一个吉洪诺夫空间. 设 $G(X)$ 是所有从 X 映射到 $[0, 1]$ 的具有 $[0, 1]$ 上的通常拓扑的连续函数所组成的集合. K 是从 $G(X)$ 映射到 $[0, 1]$ 的所有函数的集合，关于积拓扑，它是一个紧豪斯多夫空间(吉洪诺夫定理 2.2.8). 对于 $G(X)$ 中的每个 g 和 X 中的每个 x，令 $f(x)(g) := g(x)$，则 f 是一个从 X 映射到 K 的函数. 要证明 f 是连续的，(由推论 2.2.7)只需验证对积拓扑的标准子基中的每个 U，$f^{-1}(U)$ 在 X 中是开的，即对每个 $g \in G(X)$ 和 $\mathbb{R}$ 中的开集 V，所有集合 $\{y \in K: y(g) \in V\}$ 组成的集类. 对于这样的一个集合，$f^{-1}(U) = g^{-1}(V)$ 且是开的，因为 g 是连续的. 下面要证明 f 是一个同胚，首先注意到它是 1-1 的，因为对于 X 中的 $x \neq y$，由完全正则性知，存在一个 $g \in G(X)$，使得 $g(x) \neq g(y)$，因此 $f(x) \neq f(y)$. 设 W 是 X 中的任意开集，下面证明直接像 $f(W) := \{f(x): x \in W\}$ 在 $f(X)$ 中是相对开的. 设 $x \in W$，由完全正则空间的定义，在 X 上取一个连续实值函数 g，使得在 X 上，$g(x) = 1$，对所有 $y \in X \setminus W$，有 $g(y) = 0$. 不妨假设 $0 \leq g \leq 1$，用 $\max(0, \min(g, 1))$ 代替 g，则 $g \in G(X)$. 设 U 是所有 $z \in K$，满足 $z(g) > 0$ 的 z 的集合，于是 U 是开的. 满足 $f(X)$ 的 U 的交集包含在 $f(W)$ 中并且含有 $f(x)$，于是 $f(W)$ 是相对开的，且 f 是一个从 X

73

映上到$f(X)$的同胚. □

在最后的证明中，K 中值域$f(X)$的闭包是一个紧的豪斯多夫空间，称为 X 上的斯通-切赫(Stone-Čech)紧化，虽然在历史上更准确地称之为吉洪诺夫-切赫(Tychonoff-Čech)紧化.

紧化的另一种方法应用于空间 Y^J，这里的$(Y, \mathcal{T})$是一个吉洪诺夫拓扑空间，Y^J 是所有从 J 映射到 Y 的函数所组成的集合，具有积拓扑. 则如果$(K, \mathcal{U})$是$(Y, \mathcal{T})$的紧化，容易看出，K^J 关于积拓扑是 Y^J 的一个紧化. 例如，如果 $Y=\mathbb{R}$ 令 $\bar{R}$ 是两点紧致化$[-\infty, \infty]$，于是对于任意的集合 J，J 上所有实值函数所组成的空间 R^J 有一个紧致化 $\bar{R}^J$.

习题

1. 证明：(具有通常拓扑的)$\mathbb{R}^k$ 的一点紧致化与(具有$\mathbb{R}^{k+1}$的相对拓扑的)球面 $S^k:=\{(x_1, \cdots, x_{k+1})\in\mathbb{R}^{k+1}: x_1^2+\cdots+x_{k+1}^2=1\}$是同胚的.
   (a)对 $k=1$(这里 S^1 是一个圆).
   (b)对于一般的 k.[提示：在$\mathbb{R}^{k+1}$中，令 S 是半径为 1 的球面，圆心 $p=(0, \cdots, 0, 1)$，$S=\{y: |y-p|=1\}$，对于 S 中的每个 $y\neq 2p$，通过 $2p$ 和 y 的唯一直线与集合$\{x: x_{k+1}=0\}$交于唯一一点 $g(y)$]. 证明：g 给出了从 $S\setminus\{2p\}$映上到$\mathbb{R}^k$ 的同胚.
2. 设$(X, \mathcal{T})$有一个紧化(K, f)，这里的 K 仅仅包含了不在f的值域内的一个点且 K 是一个紧豪斯多夫空间. 证明：$(X, \mathcal{T})$是局部紧的.
3. 证明：对于任意的吉洪诺夫空间 X，X 上的任意有界连续实值函数都可以扩张成 X 的吉洪诺夫-切赫紧化上的函数.

74

4. 设$(X, \mathcal{T})$是一个局部紧的豪斯多夫空间，证明：X 作为它的吉洪诺夫-切赫紧化 K 的子集是开的(如果f是从 X 映射到吉洪诺夫-切赫紧化定义中给定的 K 的同胚，那么f的值域是开的). [提示：给定任意 $x\in X$，设 U 是 x 的有紧闭包的邻域. 证明在 X 上，存在一个连续的实值函数f，使得在 x 处，$f=0$，在 $X\setminus U$ 上，$f=1$. 利用这个函数证明 x 不在 $K\setminus U$ 的闭包中.]
5. 设 K 是$\mathbb{R}$ 的吉洪诺夫-切赫紧化，证明：从$\mathbb{R}\times\mathbb{R}$ 映上到$\mathbb{R}$ 的加法不能扩张成从 $K\times K$ 映射到 K 的连续函数 S 上.[提示：设$\mathbb{R}$ 中的网格 x_α，在 K 中收敛到点 $x\in K/\mathbb{R}$. 则 $-x_\alpha$ 收敛到某点 $y\in K\setminus\mathbb{R}$. 如果 S 存在，则 $S(x, y)=0\in\mathbb{R}$，于是由习题 4 知，在 K 中一定存在 X 的邻域 U 和 y 邻域 V，使得对所有的 $u\in U$ 和 $v\in V$，有 $S(u, v)\in\mathbb{R}$ 且 $|S(u, v)|<1$. 然而能证明 U 和 V 中的每个都包含绝对值任意大的实数，得到矛盾.
6. 设(X, d)是一个局部紧可分的度量空间，证明：它的一点紧化是可度量的.
7. 设 X 是任意的非紧度量空间，把它考虑作为它的吉洪诺夫-切赫紧化 K 的一个子集，令 $y\in K\setminus X$，通过证明在 K 中不存在序列 $x_n\in X$，使得 $x_n\to y$ 来证明 K 不是可度量的. [提示：如果 $x_n\to y$，假定所有的 x_n 都是不相同的，通过取一个子序列，则$\{x_{2n}\}_{n\geq 1}$与$\{x_{2n-1}\}_{n\geq 1}$构成了 X 中不相交的闭集. 利用引理 2.6.3 得到一个 X 上的连续函数f，使得对所有的 n，有 $f(x_{2n})=1$，$f(x_{2n-1})=0$，于是在 K 中$\{x_n\}$不收敛到 y.]
8. 证明：对于任意的度量空间 S，如果 $A\subset S$，$x\in\bar{A}\setminus A$，则在 A 上存在一个有界的连续实值函数，它不能被扩张成 $A\cup\{x\}$上的连续函数.[提示：对于 $t>0$，$f(t):=\sin(1/t)$不能连续地扩张到 $t=0$.]

注释

2.1 节 根据 Grattan-Guinness(1970, p. 51—53, 76)，实变量实值函数的极限和连续的定义最先是由 Bolzano(1818)提出，然后是 Cauchy(1821).

对于实数集合，Cantor(1872)给出了"邻域"和"聚点"的概念. 一个点 x 是集合 A 的聚点，当且仅当 x 的每个邻域度都包含了 A 中不同于 x 的点. 康托尔发表了一系列文章，在这些文章中，他发

展了集合 X 的“导出集”Y 的概念，这里 Y 是 X 的所有聚点的集合，此结论对于多维情形也成立(Cantor，1879—1883)．在这些文章中还涉及了开集、闭集、内点和闭包的概念，至少隐含了这些

75

概念．(一个集合的闭包是含有它的第一个导出集的并)．Maurice Fréchet(1906，p. 17，30)开始了度量空间的研究．Siegmund-Schultze(1982，第 4 章)研究了相关的历史．弗雷歇(Fréchet)给出了没有距离的序列收敛的抽象公式．豪斯多夫 Hausdorff(1914，第 9 章的第 1 节)在定义了豪斯多夫空间之后，借助于开集给出了连续函数的定义．Kuratowski(1958，p. 20，29)回顾了这些概念和其他一些概念在拓扑空间、闭包等方面的贡献．网格及其收敛的概念要归于 E. H. Moore(1915)，H. L. Smith 也部分地参与了此工作(Moore 和 Smith，1922)．Henri Cartan(1937)定义了滤子和超滤子．Caratheodory(1913，p. 331)更早地研究了非空集合的递减序列，它可以被看作是滤子基，M. H. Stone(1936) 定义了“对偶理想”，不包含∅的对偶理想，则是滤子.

Hausdorff(1914) 在一般拓扑而不必是度量空间方面出版了第一本著作，也首次证明了关于度量空间的基本事实(参见 2.3 节、2.5 节的注释)．豪斯多夫生于 1868 年卒于 1942 年．正如 Eichhorn(1992)所言，豪斯多夫一生著作甚多，覆盖各类文学和哲学，其中有以笔名 Paul Mongré 发表的诗歌和一部戏剧．由于是犹太人，从 1933 年开始，在纳粹统治下，他偶遭不幸．1942 年，豪斯多夫与其夫人和妹妹一起过着逃离生活，未被送到集中营.

Heine(1872，p. 186)通过连续的区间对分法证明了闭区间上的连续实值函数有最大值和最小值，这是在定理 2.1.2 的证明之后作为一个例子给出的．Fréchet(1918)首创了 L^*-空间．Kisyński(1959—1960)证明了如果 C 是 L^*-收敛的，则 $C(T(C))=C$.

Alexandroff 和 Fedorchuk (1978) 总结了集合论拓扑的历史，给出了 369 篇参考文献．参见 Arboleda(1979).

2.2 节 Bourbaki(1953，p. 45)这本重要的著作包括了紧拓扑空间上的豪斯多夫分离条件(“séparé”)．而大多数人更喜欢写成紧豪斯多夫空间．一些与紧性相关的概念最初是以序列、可数开覆盖等形式给出的，只是后来才给出较一般的形式．Bernard Bolzano(1781—1848)给出了其中一个紧性概念，主要结论为每个有界的无限实数集合都有一个聚点．据 van Rootselaar(1970)所言，在波尔查诺(Bolzano)的书里并没有此结论的证明，在奥地利国家图书馆里保存的大部分都是未出版的原稿．很明显，波尔查诺在某些方面犯了一些错误.

Borel(1895，p. 51—52)证明 $\mathbb{R}$ 中任意覆盖有界闭区间的开区间序列都有一个有限子覆盖．Lebesgue(1904，p. 117)将这个定理拓展到由任意二维集合的开区域(同胚于一个开圆盘)覆盖，这个二维集合是[0，1]上的连续像，尤其是通过佩亚诺曲线，任意集合同胚于一个闭正方形的集合．Lebesgue(1907b)和 Temple(1981)研究了海涅–博雷尔或海涅–博雷尔–勒贝格定理的发展历程．Borel(1895)首次明确给出了这个结论，海涅给出了明确的证明.

一般地，紧空间目前的定义要归功于 Alexandroff 和 Urysohn(1924)．Alexandroff(1926，p. 561)证明了紧空间上连续函数的值域是紧的.

Fréchet(1906)证明了[0，1]的同步可数积是紧的．Tychonoff(1929—1930)也证明了[0，1]的

76

同步任意积是紧的．Čech(1937)证明了吉洪诺夫定理，即紧空间的任意笛卡儿积是紧的，根据 Kelley(1955，p. 143)，这可能是一般拓扑中最重要的定理了．Eduard Čech 生于 1893 年卒于 1960 年，人们收集了他关于拓扑方面的论文(Čech，1968)．《Point Sets》(Čech 1936，1966，1969)和《Topological Spaces》(Čech 1959，1966)是关于一般拓扑的著作，它们在切赫去世后均被翻译成英

文，但是紧空间的积在这两部著作中却没有提到．在Čech(1968)的绪论中给出10页的科学传记，M. Katětov、J. Novák和A. Švec介绍了切赫的生活和工作．Čech的贡献主要集中在代数学和一般拓扑上.

Alexandroff et al. (1956，1967)为祝贺吉洪诺夫的50和60岁生日特意写了些文章．事实上，吉洪诺夫的研究主要集中在微分方程和数学物理方面.

H. Cartan(1937)定义了超滤子，并证明了拓扑空间是紧的当且仅当每个超滤子都是收敛的(定理2.2.5)．他还证明了对集合X中的任意超滤子和从X到集合Y的函数f，$\mathcal{U}$的直接象(即$\{B\subset Y: f^{-1}(B)\in\mathcal{U}\}$)是个超滤子．虽然Cartan并没有明确提出吉洪诺夫定理，且在同一年切赫一般形式首次出现，但从这些事实不难得到吉洪诺夫定理的超滤子证法．Bourbaki(1940)给了一种证明，Kelley(1955)给了两种证明．参考Bourbaki，第二种证法接近于超滤子证法，但却没有明确指出滤子和超滤子．Kelley的书的第2章是关于网格的(Moore-Smith收敛)，但只是在这章末的问题L中提到了滤子．也可参考Chernoff(1992)对吉洪诺夫定理的证明．Feferman(1964，定理4.12，p.343)证明了在没有选择公理的情况下，它和集合论是一致的，在N中集合论就是：仅有的超滤子是点超滤子.

Alexandroff(1923，p.561)证明了定理2.2.11. 他也在代数拓扑上展开过研究，例如，他首次给出了抽象序列的概念. Pontryagin和Mishchenko(1956)、Kolmogoroff et al. (1966)，以及Arkhangelskii et al. (1976)都曾写文章来祝贺Alexandroff的60、70和80岁生日.

2.3节 Hausdorff(1914，p.311—315)第一次给出了完全有界的概念，接着证明了度量空间是紧的，当且仅当它既是完全有界的又是完备的.

2.4节 以发现积分定理和复分析中的积分公式出名的柯西在1823年错误地断言：如果连续函数序列在一个区间的每个点处都收敛，那么它们的和在这个区间上是连续的. Abel(1826)给出一个反例$\sum_n(-1)^n(\sin nx)/n$，它在$0\leqslant x<\pi$处收敛到$x/2$，在π处收敛到0.(阿贝尔生于1802年卒于1829年)．Cauchy(1833，p.55—56)没有注意到这个问题，且又重复了这个错误．特殊情况下的一致收敛概念第一次是出现在Abel(1826)和Gudermann(1838，p.251—252)的论文中．Manning(1975，p.363)写到"1853年以前的所有涉及到逐项幂级数积分的柯西证明都是错误的，因为这都用到了一致收敛这个概念."连续函数的一致极限是连续的这个定理(2.4.8中是实变量函数)，分别由Seidel(1847—1849)和Stokes(1847—1848)所证明．斯托克(Stokes)错误地断言：如果连续函数序列在闭的有界区间上逐点收敛到一个连续函数，则收敛一定是一致的. Seidel注意到他不能给出相反结论的证明．伟大的数学家hardy(1916—1919)在稍后的时期内检验了斯托克及其他人的贡献.最终，Cauchy(1853)将柯西序列的概念用公式表示了出来，并证明了$\mathbb{R}$(命题2.4.2)和C_b(定理2.4.9，是对有界集合上的实函数和复变量函数)的完备性.

77

Heine(1870，p.361)是定义函数一致连续性的第一个人，并且(1872，p.188)证明了任意有界闭区间上的实值连续函数都是一致连续的(推论2.4.6的原型)．海涅极信任魏尔斯特拉斯未出版的文稿．虽然魏尔斯特拉斯在其他领域的著作在其死后陆续出版，但是根据Biermann(1976)所言，他在实函数方面的著作却一直未出版．海涅出生在1821年，全名为Heinrich Eduard Heine，他的著作都以Eduard为名出版，大概是为了与著名诗人Heinrich Heine(1797—1856)相区分吧．根据首创$d_{\sup}$度量(或一般度量)的弗雷歇所言，魏尔斯特拉斯是第一个系统地应用一致收敛的数学家(参考Manning，1975)．等度连续的概念归功于Arzelà(1882—1883)和Ascoli(1883—1884)．尽管Dini

(1878)已经较早地证明了与阿尔泽拉-阿斯科利定理(2.4.7)相关的结论，这点 Dunford 和 Schwartz (1958, p.382—383) 给予了标注，但是这个定理却是在 Ascoli(1883—1884)和 Arzelà(1889, 1895)的论文中完成的. 迪尼以其在实分析中的单调收敛定理(2.4.10)而闻名. 他还在微分几何方面做过很重要的研究(Dini, 1953, 卷 1; Reich, 1973). Baire(1906)研究得到$\mathbb{R}$与$(-1, 1)$同胚. Hahn (1921)证明了任意度量空间都同胚于一个有界的度量空间(命题 2.4.3). Fréchet(1928)度量了可数个度量空间的积(命题 2.4.4).

M. H. Stone(1947—1948)证明了斯通-魏尔斯特拉斯定理 2.4.11. Weierstrass(1885, p.5, 36)分别证明了 $d=1$ 和 d 为任意有限值时的多项式逼近定理(推论 2.4.12). 魏尔斯特拉斯的卷积法看起来需要定义在紧集 K 的一个邻域上的连续函数，这一点是可以做到的，例如，通过扩张定理 2.6.4 即可得到. 在$\mathbb{R}$内的一个有界闭区间上，伯恩斯坦多项式给出了一个精确的逼近，例如，参考 Bartle(1964, 定理 17.6).

Peano(1890)定义了值域为一个正方形的曲面(习题 2.4.9).

2.5 节 Hausdorff(1914, p.315—316)证明了每个度量空间都有一个完备的度量空间(定理 2.5.1). Kuratowski(1935, p.543)给出了对有界空间的简短证明. 之后不久人们就注意到并将之扩张到无界空间上. 我要感谢 L. Š. Grinblat，是他告诉了我这个证明. Kuratowski 的证明思想和 Fréchet (1910, p.159—161)的证明思想有某些关联. 在很多著作中，Kuratowski 的证明往往是沿着如下提纲进行的：对 S 中任意两个柯西序列$\{x_n\}$和$\{y_n\}$，令 $e(\{x_n\}, \{y_n\}) := \lim_{n\to\infty} d(x_n, y_n)$. 要证明这个极限是存在的且它定义了 S 中所有柯西序列集合上的伪度量. 定义一个关系 E，使得 xEy 当且仅当 $e(x, y)=0$. 这是一个等价关系. 在关于 E 的所有等价类的集合 T 上，e 定义了一个度量. 令 f 是一个从 S 映射到 T 的函数，满足对每个 S 中的 x，$f(x)$是柯西序列$\{x_n\}$的等价类，其中对所有的 n，$x_n = x$. 因此，(T, e)是 X 的完备化. 尽管证明看起来很自然，概念上也很直接，但是在具体的证明中却涉及很多细节方面的东西.

78

范畴定理(2.5.2)通常称为贝尔(Baire)范畴定理. 事实上，Osgood(1897, p.171—173)最初是在$\mathbb{R}$内证明的. 然后 Baire(1899, p.65)又证明了对$\mathbb{R}^n$也是成立的. Mazurkiewicz(1916)证明了每个拓扑上的完备空间都是其完备化中的一个 G_δ. Dugundji(1966)给出了逆结论(定理 2.5.4). Alexandroff (1924)也证明了这个逆定理，Hausdorff(1924)给出了更简短的证明.

2.6 节 Lebesgue(1907a)证明了对$\mathbb{R}^2$的扩张定理 2.6.4，他并没有直接扩张到$\mathbb{R}^k$. Tietze(1915, p.14)第一次证明了 f 有界且 X 是度量空间条件下的定理. Borsuk(1934, p.4)证明了如果闭子集 F 在度量空间 X 中是可分的，那么从 $C_b(F)$映射到 $C_b(X)$的有界实值连续函数的扩张可以选择线性的. 另一方面，Dugundji(1951)将 Tietze 定理拓展到更一般(可能是无限维值域空间)的情况.

Tietze(1923, p.301, 公理(h))定义了正规空间. 对于正规空间，Urysohn(1925, p.290—291)在他的遗稿中证明了引理 2.6.3 和扩张定理 2.6.4 中的情况(a)，给出了本质上的证明(p.293). Urysohn 生于 1898 年. Arkhangelskii et al. (1976)写到 1924 年在波恩访问期间，"Aleksandrov 和 Uryson 每天都在莱因河中游泳，他们的游泳技术并不好，这引起了豪斯多夫的不满……1924 年 8 月 17 日，年仅 26 岁的 Uryson 在大西洋中溺水身亡."关于 Uryson 的生平参考 Alexandroff(1950).

Alexandroff 和 Hopf(1935, p.76)陈述了一般情况的定理 2.6.4(情况(c))，但只证明了情况(a). Caratheodory(1918, 1927, p.619)注意到，对于 $X=\mathbb{R}^k$，可以通过 $1+d(x, F)$来分割 g 的方法由(a)可以推出(b). 这在任意的度量空间上都是有效的. 对于一般的正规空间，关于情况(b)的

补充证明，我能给出的最早文献是 Bourbaki(1948).

2.7 节 André Weil(1937)开创了一致空间的理论．比这里更进一步的探究可以参阅 Kelly(1955，第 6 章)．Kelly(p.186)将度量化定理 2.7.1 归功于 Alexandroff 和 Urysohn(1923)及 Chittenden(1927)，而它目前的公式和证明则归于 Weil(1937)、Frink(1937)和 Bourbaki(1948).

2.8 节 在一点紧化方面 P. S. Alexandroff 做出了卓越贡献．一点紧化是作为例子出现在 Alexandroff 和 Hopf(1935，p.93)的论文中．正规豪斯多夫空间又称为 T_4 空间．如果对每个不在闭集 F 中的点 p，存在不相交的开集 U 和 V，使得 $p\in U$，$F\subset V$，则称$(X,\mathcal{T})$是正则的．豪斯多夫正则空间又称为 T_3 空间．显然，每个 T_4 空间也是 $T_{3.5}$(＝吉洪诺夫)空间，而且每个吉洪诺夫空间也是 T_3 空间．Urysohn(1925，p.292)使用了完全正则的假设，但却尚未对它命名.

Tychonoff(1930)证明了每个 T_4 空间都有一个紧化，定义了完全正则空间，并给出了是 T_3 空间而不是 $T_{3.5}$空间，是 $T_{3.5}$空间而不是 T_4 空间的例子．他还证明豪斯多夫空间有一个(豪斯多夫)紧化，当且仅当它是完全正则的(定理 2.8.3).

论文Čech(1937)的第一段明确写到 Tychonoff(1930)证明了对一个紧豪斯多夫空间 $\beta(S)$ 的任意完全正则($T_{3.5}$)空间 S，存在(i)S 同胚于 $\beta(S)$ 的一个稠密子集，(ii)每个有界连续实函数可扩张成 $\beta(S)$ 上的函数．切赫描述“很容易看到 $\beta(S)$ 可以由性质(i)和(ii)唯一确定．此论文的目标主要是研究 $\beta(S)$．”这样，“Stone-Čech”紧化 $\beta(S)$ 要归功于吉洪诺夫，并且切赫对此作了更进一步的研究．Stone(1937，p.455ff.，esp.461—463)用了很大的篇幅来讨论紧化，这个关系中引用吉洪诺夫只是为了说明 $T_4\to T_{3.5}\to T_3$ 的逆不成立. [79]

参考文献

Lebesgue(1907b)修订 W. H. Young 及 G. C. Young(1906)在点集和拓扑方面 300 余篇文献构成的“bibliographie trop copieuse”，称其为分析拓扑学．这里给出 90 个参考文献．带星号的资料在次级来源中被讨论过，在原著中并没见到. [80]

Abel, Niels Henrik (1826). Untersuchungen über die Reihe $1+\frac{m}{1}x+\frac{m(m-1)}{1\cdot 2}x^2+\frac{m(m-1)(m-2)}{1\cdot 2\cdot 3}x^2+\cdots$, *J. reine angew. Math. 1*: 311–339. Also in *Oeuvres complètes*, ed. L. Sylow and S. Lie. Grøndahl, Kristiania [Oslo], 1881, I, pp. 219–250.

Alexandroff, Paul [Aleksandrov, Pavel Sergeevich] (1924). Sur les ensembles de la première classe et les espaces abstraits. *Comptes Rendus Acad. Sci. Paris 178*: 185–187.

——— (1926). Über stetige Abbildungen kompakter Räume. *Math. Annalen 96*: 555–571.

——— (1950). Pavel Samuilovich Urysohn. *Uspekhi Mat. Nauk 5*, no. 1: 196–202.

——— et al. (1967). Andrei Nikolaevich Tychonov (on his sixtieth birthday): On the works of A. N. Tychonov in .... *Uspekhi Mat. Nauk 22*, no. 2: 133–188 (in Russian); transl. in *Russian Math. Surveys 22*, no. 2: 109–161.

——— and V. V. Fedorchuk, with the assistance of V. I. Zaitsev (1978). The main aspects in the development of set-theoretical topology. *Russian Math. Surveys 33*, no. 3: 1–53. Transl. from *Uspekhi Mat. Nauk 33*, no. 3 (201): 3–48 (in Russian).

——— and Heinz Hopf (1935). *Topologie*, 1. Band. Springer, Berlin; repr. Chelsea, New York, 1965.

———, A. Samarskii, and A. Sveshnikov (1956). Andrei Nikolaevich Tychonov (on his fiftieth birthday). *Uspekhi Mat. Nauk 11*, no. 6: 235–245 (in Russian).

——— and Paul Urysohn (1923). Une condition nécessaire et suffisante pour qu'une classe ($\mathcal{L}$) soit une classe ($\mathcal{D}$). *C. R. Acad. Sci Paris 177*: 1274–1276.

——— and——— (1924). Theorie der topologischen Räume. *Math. Annalen. 92*: 258–266.

Arboleda, L. C. (1979). Les débuts de l'école topologique soviétique: Notes sur les lettres de Paul S. Alexandroff et Paul S. Urysohn à Maurice Fréchet. *Arch. Hist. Exact Sci. 20*: 73–89.

Arkhangelskii, A. V., A. N. Kolmogorov, A. A. Maltsev, and O. A. Oleinik (1976). Pavel Sergeevich Aleksandrov (On his eightieth birthday). *Uspekhi Mat. Nauk 31*, no. 5: 3–15 (in Russian); transl. in *Russian Math. Surveys 31*, no. 5: 1–13.

*Arzelà, Cesare (1882–1883). Un'osservazione intorno alle serie di funzioni. *Rend. dell' Acad. R. delle Sci. dell'Istituto di Bologna*, pp. 142–159.

*——— (1889). Funzioni di linee. *Atti della R. Accad. dei Lincei, Rend. Cl. Sci. Fis. Mat. Nat.* (Ser. 4) *5*: 342–348.

*——— (1895). Sulle funzioni di linee. *Mem. Accad. Sci. Ist. Bologna Cl. Sci. Fis. Mat.* (Ser. 5) *5*: 55–74.

*Ascoli, G. (1883–1884). Le curve limiti di una varietà data di curve. *Atti della R. Accad. dei Lincei, Memorie della Cl. Sci. Fis. Mat. Nat.* (Ser. 3) *18*: 521–586.

Baire, René (1899). Sur les fonctions de variables réelles (Thèse). *Annali di Matematica Pura ed Applic.* (Ser. 3) *3*: 1–123.

——— (1906). Sur la représentation des fonctions continues. *Acta Math. 30*: 1–48.

Bartle, R. G. (1964). *The Elements of Real Analysis*. Wiley, New York.

Biermann, Kurt-R. (1976). Weierstrass, Karl Theodor Wilhelm. *Dictionary of Scientific Biography 14*: 219–224. Scribner's, New York.

*Bolzano, B. P. J. N. (1818). Rein analytischer Beweis des Lehrsatzës, dass zwischen je zwey Werthen, die ein entgegengesetztes Resultat gewähren, wenigstens eine reelle Wurzel der Gleichung liege. *Abh. Gesell. Wiss. Prague* (Ser. 3) *5*: 1–60.

Borel, Émile (1895). Sur quelques points de la théorie des fonctions. *Ann. Scient. Ecole Normale Sup.* (Ser. 3) *12*: 9–55.

Borsuk, Karol (1934). Über Isomorphie der Funktionairäume. *Bull. Acad Polon. Sci. Lett. Classe Sci. Sér. A Math.* 1933: 1–10.

Bourbaki, Nicolas [pseud.] (1940, 1948, 1953, 1961, 1971). *Topologie Générale, Chap. 9. Utilisation des nombres réels en topologie générale*. Hermann, Paris. English transl. *Elements of Mathematics. General Topology, Part 2*. Hermann, Paris; Addison-Wesley, Reading, Mass., 1966.

Cantor, Georg (1872). Ueber die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen. *Math. Annalen 5*: 123–132.

——— (1879–1883). Ueber unendliche, lineare Punktmannichfaltigkeiten. *Math. Annalen 15*: 1–7; *17*: 355–358; *20*: 113–121; *21*: 51–58.

Caratheodory, Constantin (1913). Über die Begrenzung einfach zusammenhängender Gebiete. *Math. Annalen 73*: 323–370.

——— (1918). *Vorlesungen über reelle Funktionen*. Teubner, Leipzig and Berlin. 2d ed., 1927.

Cartan, Henri (1937). Théorie des filtres: Filtres et ultrafiltres. *C. R. Acad. Sci. Paris 205*: 595–598, 777–779.

Cauchy, Augustin-Louis (1821). *Cours d'analyse de l'École Royale Polytechnique*, Paris.

——— (1823). *Résumés des leçons données à l'École Royale Polytechnique sur le calcul infinitésimal*. Debure, Paris; also in *Oeuvres Complètes* (Ser. 2), IV, pp. 5–261, Gauthier-Villars, Paris, 1889.

*——— (1833). *Résumés analytiques*. Imprimerie Royale, Turin.

——— (1853). Note sur les séries convergentes dons les divers termes sont des functions continues d'une variable réelle ou imaginaire, entre des limites données. *Comptes Rendus Acad. Sci. Paris 36*: 454–459. Also in *Oeuvres Complètes* (Ser. 1) XII, pp. 30–36. Gauthier-Villars, Paris, 1900.

Čech, Eduard (1936). *Point Sets*. In Czech, *Bodové Množiny*. 2d ed. 1966, Czech. Acad. Sci., Prague; *In English, transl. by Aleš Pultr, Academic Press, New York, 1969.

——— (1937). On bicompact spaces. *Ann. Math.* (Ser. 2) *38*: 823–844.

——— (1959). *Topological Spaces*. In Czech, **Topologické Prostory*. Rev. English Ed. (1966), Eds. Z. Frolík and M. Katětov; Czech. Acad. Sci., Prague; Wiley, London.

——— (1968). *Topological Papers of Eduard Čech*. Academia (Czech. Acad. Sci), Prague.

Chernoff, Paul R. (1992). A simple proof of Tychonoff's theorem via nets. *Amer. Math. Monthly 99*: 932–934.

Chittenden, E. W. (1927). On the metrization problem and related problems in the theory of abstract sets. *Bull. Amer. Math. Soc. 33*: 13–34.

*Dini, Ulisse (1878). *Fondamenti per la teorica delle funzioni di variabili reali*. Nistri, Pisa.

——— (1953–1959, posth.) *Opere. 5* vols. Ediz. Cremonese, Rome.

Dugundji, James (1951). An extension of Tietze's theorem. *Pacific J. Math. 1*: 352–367.

——— (1966). *Topology*. Allyn & Bacon, Boston.

Dunford, Nelson, and Jacob T. Schwartz, with the assistance of William G. Badé and Robert G. Bartle (1958). *Linear Operators, Part I: General Theory*. Interscience, New York.

Eichhorn, Eugen (1992). Felix Hausdorff—Paul Mongré. Some aspects of his life and the meaning of his death. In Gähler, W., Herrlich, H., and Preuß, G., eds., *Recent Developments of General Topology and its Applications; Internat. Conf. in Memory of Felix Hausdorff (1868–1942)*, Akademie Verlag, Berlin, pp. 85–117.

Feferman, Solomon (1964). Some applications of the notions of forcing and generic sets. *Fund. Math. 56*: 325–345.

Fréchet, Maurice (1906). Sur quelques points du calcul fonctionnel. *Rendiconti Circolo Mat. Palermo* 22: 1–74.

——— (1910). Les dimensions d'un ensemble abstrait. *Math. Annalen 68*: 145–168.

——— (1918). Sur la notion de voisinage dans les ensembles abstraits. *Bull. Sci. Math. 42*: 138–156.

——— (1928). *Les espaces abstraits*. Gauthier-Villars, Paris.

Frink, A. H. (1937). Distance functions and the metrization problem. *Bull. Amer. Math. Soc. 43*: 133–142.

Grattan-Guinness, I. (1970). *The Development of the Foundations of Analysis from Euler to Riemann*. MIT Press, Cambridge, MA.

Gudermann, Christof J. (1838). Theorie der Modular-Functionen und der Modular-Integrale, 4–5 Abschnitt. *J. reine angew. Math. 18*: 220–258.

Hahn, Hans (1921). *Theorie der reellen Funktionen 1*. Springer, Berlin.

Hardy, Godfrey H. (1916–1919). Sir George Stokes and the concept of uniform convergence. *Proc. Cambr. Phil. Soc. 19*: 148–156.

Hausdorff, Felix (1914). *Grundzüge der Mengenlehre*. Von Veit, Leipzig. (See References to Chap. 1 for later editions.)

——— (1924). Die Mengen G_δ in vollständigen Räumen. *Fund. Math. 6*: 146–148.

Heine, E. [Heinrich Eduard] (1870). Ueber trigonometrische Reihen. *Journal für die reine und angew. Math. 71*: 353–365.

——— (1872). Die Elemente der Functionenlehre. *Journal für die reine und angew. Math. 74*: 172–188.

Hoffman, Kenneth M. (1975). *Analysis in Euclidean Space*. Prentice-Hall, Englewood Cliffs, N.J.

Kelley, John L. (1955). *General Topology*. Van Nostrand, Princeton.

Kisyński, J. (1959–1960). Convergence du type L. *Colloq. Math. 7*: 205–211.

Kolmogorov, A. N., L. A. Lyusternik, Yu. M. Smirnov, A. N. Tychonov, and S. V. Fomin (1966). Pavel Sergeevich Alexandrov (On his seventieth birthday and the fiftieth anniversary of his scientific activity). *Uspekhi Mat. Nauk 21*, no. 4: 2–7 (in Russian); transl. in *Russian Math. Surveys 21*, no. 4: 2–6.

Kuratowski, Kazimierz (1935). Quelques problèmes concernant les espaces métriques non-séparables. *Fund. Math. 25*: 534–545.

——— (1958). *Topologie*, vol. 1. PWN, Warsaw; Hafner, New York.

Lebesgue, Henri (1904). *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthier-Villars, Paris.

——— (1907a). Sur le problème de Dirichlet. *Rend. Circolo Mat. Palermo 24*: 371–402.

——— (1907b). Review of Young and Young (1906). *Bull. Sci. Math.* (Ser. 2) *31*: 129–135.

Manning, Kenneth R. (1975). The emergence of the Weierstrassian approach to complex analysis. *Arch. Hist. Exact Sci. 14*: 297–383.

*Mazurkiewicz, Stefan (1916). Über Borelsche Mengen. *Bull. Acad. Cracovie 1916:* 490–494.

Moore, E. H. (1915). Definition of limit in general integral analysis. *Proc. Nat. Acad. Sci. USA 1*: 628–632.

——— and H. L. Smith (1922). A general theory of limits. *Amer. J. Math. 44*: 102–121.

Osgood, W. F. (1897). Non-uniform convergence and the integration of series term by term. *Amer. J. Math. 19*: 155–190.

Peano, G. (1890). Sur un courbe, qui remplit toute une aire plane. *Math. Ann. 36*: 157–160.

Pontryagin, L. S., and E. F. Mishchenko (1956). Pavel Sergeevich Aleksandrov (On his sixtieth birthday and fortieth year of scientific activity). *Uspekhi Mat. Nauk 11*, no. 4: 183–192 (in Russian).

Reich, Karin (1973). Die Geschichte der Differentialgeometrie von Gauss bis Riemann (1828–1868). *Arch. Hist. Exact Sci. 11*: 273–382.

van Rootselaar, B. (1970). Bolzano, Bernard. *Dictionary of Scientific Biography*, II, pp. 273–279. Scribner's, New York.

*Seidel, Phillip Ludwig von (1847–1849). Note über eine Eigenschaft der Reihen, welche discontinuierliche Functionen darstellen. *Abh. der Bayer. Akad. der Wiss.* (Munich) *5*: 379–393.

Siegmund-Schultze, Reinhard (1982). Die Anfänge der Funktionalanalysis und ihr Platz im Umwälzungsprozess der Mathematik um 1900. *Arch. Hist. Exact Sci. 26*: 13–71.

Stokes, George G. (1847–1848). On the critical values of periodic series. *Trans. Cambr. Phil. Soc. 8*: 533–583; *Mathematical and Physical Papers I*: 236–313.

Stone, Marshall Harvey (1936). The theory of representations for Boolean algebras. *Trans. Amer. Math. Soc. 40*: 37–111.

——— (1937). Applications of the theory of Boolean rings to general topology. *Trans. Amer. Math. Soc. 41*: 375–481.

——— (1947–1948). The generalized Weierstrass approximation theorem. *Math. Mag. 21*: 167–184, 237–254. Repr. in *Studies in Modern Analysis 1*, ed. R. C. Buck, Math. Assoc. of Amer., 1962, pp. 30–87.

Temple, George (1981). *100 Years of Mathematics*. Springer, New York.

Tietze, Heinrich (1915). Über Funktionen, die auf einer abgeschlossenen Menge stetig sind. *J. reine angew. Math. 145*: 9–14.

——— (1923). Beiträge zur allgemeinen Topologie. I. Axiome für verschiedene Fassungen des Umgebungsbegriffs. *Math. Annalen 88*: 290–312.

Tychonoff, A. [Tikhonov, Andrei Nikolaevich] (1930). Über die topologische Erweiterung von Räumen. *Math. Ann. 102*: 544–561.

——— (1935). Über einen Funktionenraum. *Math. Ann. 111*: 762–766.

Urysohn, Paul [Urison, Pavel Samuilovich] (1925). Über die Mächtigkeit der zusammenhängenden Mengen. *Math. Ann. 94*: 262–295.

Weierstrass, K. (1885). Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen reeller Argumente. *Sitzungsber. königl. preussischen Akad. Wissenschaften* 633–639, 789–805. *Mathematische Werke*, Mayer & Müller, Berlin, 1894–1927, vol. 3, pp. 1–37.

Weil, André (1937). *Sur les espaces à structure uniforme et sur la topologie générale. Actualités Scientifiques et Industrielles 551*, Paris.

Young, W. H., and Grace Chisholm Young (1906). *The Theory of Sets of Points.* Cambridge Univ. Press.



第3章 测 度

3.1 测度初步

区间长度是关于测度的一个典型例子．现代测度论是由博雷尔和勒贝格大约在1900年发展起来的，首要任务就是把长度的概念拓展到实直线的一般子集上．在把区间表示为其他区间的有限不相交的并时，用左开右闭的区间是很方便的．对 $a\leqslant b$，长度是通过 $\lambda((a,b]):=b-a$ 来定义的．在扩充的实数体系 $[-\infty,\infty]:=\{-\infty\}\cup\mathbb{R}\cup\{+\infty\}$ 中，$-\infty$ 和 $+\infty$ 不是两个真正的实数．$+\infty$ 通常简单地记作 ∞．对任意实数 x，令 $-\infty<x<\infty$，从而拓展了实数的线性顺序．正如在2.2节中所定义的，区间拓扑可以收敛到 $\pm\infty$．例如，$x_n\to+\infty$ 当且仅当对任意的 $K<\infty$，存在一个 m，对所有的 $n>m$，$x_n>K$. 如果一个实数序列或实级数称为收敛的，但是它的极限不是具体的，这时假定其极限在$\mathbb{R}$中，但不是 $\pm\infty$．对任意实数 x，$x+(-\infty):=-\infty$，$x+\infty:=+\infty$，而 $\infty-\infty$ 或 $\infty+(-\infty)$ 没有意义，尽管当 a_n+b_n 趋于一个有限极限时，可能有 $a_n\to+\infty$ 和 $b_n\to-\infty$．

令 X 是一个集合，$\mathcal{C}$是 X 的子集构成的集类且 $\varnothing\in\mathcal{C}$. 回想集合序列 $A_n(n=1,2,\cdots)$ 称为是不相交的，当且仅当 $i\neq j$ 时，$A_i\cap A_j=\varnothing$. 由$\mathcal{C}$映射到 $[-\infty,\infty]$ 的函数 μ 称为是有限可加的(finitely additive)，当且仅当 $\mu(\varnothing)=0$，且对所有的 $i=1,2,\cdots,n$，A_i 互不相交时，$A_i\in\mathcal{C}$，且

$$A:=\bigcup_{i=1}^{n}A_i\in\mathcal{C},\quad 就有\ \mu(A)=\sum_{i=1}^{n}\mu(A_i).$$

(因此，所有这种和必须有定义，以免对某个 i 和 j，有 $\mu(A_i)=-\infty$ 及 $\mu(A_j)=+\infty$.) 如果 $A_n\in\mathcal{C}$，$n=1,2,\cdots$，A_n 是互不相交的且 $B:=\bigcup_{n\geqslant1}A_n\in\mathcal{C}$，有 $\mu(B)=\sum_{n\geqslant1}\mu(A_n)$，那么称 μ 为可数可加的(countably additive).

回想对任意集合 X，幂集 2^X 是由 X 的所有子集组成的集类．

例：在集合 X 中令 $p\neq q$，如果 A 包含 p 和 q，令 $m(A)=1$，否则 $m(A)=0$，那么 m 在集类 2^X 上不是可加的．

定义 给定一个集合 X，集类 $\mathcal{A}\subset2^X$ 称为环(ring)，当且仅当 $\varnothing\in\mathcal{A}$ 且对所有的 $A,B\in\mathcal{A}$，有 $A\cup B\subset\mathcal{A}$，$B\setminus A\in\mathcal{A}$. 一个环$\mathcal{A}$称为**代数**当且仅当 $X\in\mathcal{A}$. 一个代数$\mathcal{A}$称为是**σ-代数**(σ-algebra)，如果对任意的集合序列 $\{A_n\}\subset\mathcal{A}$，有 $\bigcup_{n\geqslant1}A_n\in\mathcal{A}$.

例如，在任意集合 X 中，所有有限集组成的集类是一个环，但不是一个代数，除非 X 是有限的．所有有限集及其补集组成的集类是一个代数，但不是一个 σ-代数，同样除非 X 是有限的．

注意到，对环$\mathcal{R}$中的任意 A，B，$A\cap B=A\setminus(A\setminus B)\in\mathcal{R}$. 对任意集合 X，2^X 是 X 的所有子集的一个 σ-代数．对任意集类$\mathcal{C}\subset2^X$，存在包含$\mathcal{C}$的最小代数，即所有包含$\mathcal{C}$的代数的交．类似地，存在包含$\mathcal{C}$的最小 σ-代数．这些代数和 σ-代数均称为是由$\mathcal{C}$生成的．

例如，如果$\mathcal{A}$是集合 X 的所有单元素集 $\{x\}$ 组成的集类，由$\mathcal{A}$生成的代数是由 X 的所有有限子集或有有限补 $X\setminus A$ 的子集组成的集类．由 A 生成的 σ-代数是可数集或有可数补的集合组成的集类．

下面是判断可数加的第一个准则．对任意集合序列 $A_1,A_2,\cdots,A_n\downarrow\varnothing$ 意味着对所有的 n，$A_n\supset A_{n+1}$ 且 $\bigcap_n A_n=\varnothing$. 对一个无限区间，比如 $[c,\infty)$，其中 c 有限，有 $\lambda((c,\infty]):=\infty$. 那么对

$A_n := [n, \infty)$，有 $A_n \downarrow \varnothing$，但对所有的 n，$\lambda(A_n) = +\infty$，不收敛于 0. 这说明了为什么在下面的定理中要求 μ 具有实(有限)值.

3.1.1 定理　令 μ 是代数 $\mathcal{A}$ 上的一个有限可加实值函数，那么 μ 是可数可加的，当且仅当 μ 在 $\varnothing$ 上连续，即当 $A_n \in \mathcal{A}$ 且 $A_n \downarrow \varnothing$ 时，$\mu(A_n) \to 0$.

证明　首先假设 μ 是可数可加的，$A_n \in \mathcal{A}$ 且 $A_n \downarrow \varnothing$. 那么对所有的 n，集合 $A_n \setminus A_{n+1}$ 是互不相交的且其并是 A_1. 同样，对 $n \geqslant m$ 及每个 m，其并是 A_m. 对每个 m，还满足 $\sum_{n \geqslant m} \mu(A_n \setminus A_{n+1}) = \mu(A_m)$. 因为级数 $\sum_{n \geqslant 1} \mu(A_n \setminus A_{n+1})$ 收敛，所以对 $n \geqslant m$，其和一定趋于 0，因此 μ 在 $\varnothing$ 上连续.

86

反之，假设 μ 在 $\varnothing$ 上连续，集合 $B_j \in \mathcal{A}$ 是互不相交的且 $B := \bigcup_j B_j \in \mathcal{A}$. 令 $A_n := B \setminus \bigcup_{j<n} B_j$，那么 $A_n \in \mathcal{A}$ 且 $A_n \downarrow \varnothing$，因此 $\mu(A_n) \to 0$. 由有限可加性可得对每个 n，

$$\mu(B) = \mu(A_n) + \sum_{j<n} \mu(B_j).$$

令 $n \to \infty$，$\mu(B) = \sum_j \mu(B_j)$. □

定义　从由 X 的子集组成的 σ-代数 $\mathcal{S}$ 映射到 $[0, \infty)$ 的可数可加函数 μ 称为一个**测度**(measure). 那么 $(X, \mathcal{S}, \mu)$ 称为一个**测度空间**(measure space).

对 $A \subset X$，令 $\mu(A)$ 为有限集 A 的势，$+\infty$ 为无限集 A 的势. 那么 μ 总是一个测度，称为 X 上的计数测度(counting measure).

下面的结论对于重排非负项的和很有用的. 证明见附录 D.

3.1.2 引理　如果对所有的 $m, n \in \mathbb{N}$，满足 $a_{mn} \geqslant 0$ 且 $k \mapsto \langle m(k), n(k) \rangle$ 是由 $\mathbb{N}$ 映上到 $\mathbb{N} \times \mathbb{N}$ 的任意 1-1 函数，那么

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{mn} = \sum_{k=0}^{\infty} a_{m(k), n(k)}.$$

当证明长度是可数可加的且能拓展到更大的集类(一个 σ-代数)时，同样具有可数可加性，与此同时，如果以 $G(b) - G(a)$(其中 G 为一个合适的函数)代替区间 $(a, b]$ 的长度 $b - a$，证明可数可加性仍然成立，这是很有用的. 从 $\mathbb{R}$ 映射到自身的函数 G 称为是非递减的(nondecreasing)，当且仅当 $x \leqslant y$ 时，$G(x) \leqslant G(y)$. 那么对任意的 x，极限 $G(x^+) := \lim_{y \downarrow x} G(y)$ 存在. 称 G 是右连续的(continuous from the right)，当且仅当对所有的 x，$G(x^+) = G(x)$. 令 G 是一个非递减的右连续函数，则当 $x \uparrow = +\infty$ 时，对某个 $z \leqslant +\infty$，有 $G(x) \uparrow z$. 令 $G(+\infty) := z$，类似地定义 $G(-\infty)$，使得当 $x \downarrow -\infty$ 时，$G(x) \downarrow G(-\infty)$.

例如，如果 G 是区间 $(1, \infty)$ 的示性函数，那么 G 是非递减的，$G(1) = 0$ 且 $G(1^+) = 1$. 因此，G 不是右连续的(它是左连续的).

令 $\mathcal{C} := \{(a, b]: -\infty < a \leqslant b < +\infty\}$. $\mu := \mu_G$ 是定义在 $\mathcal{C}$ 上的一个函数，满足 $\mu((a, b]) := G(b) - G(a)$. 如果 $a = b$ 使得 $(a, b] = \varnothing$，则有 $\mu(\varnothing) = 0$. 那么对每个 $A \in \mathcal{C}$，$0 \leqslant \mu(A) < +\infty$.

3.1.3 定理　如果 G 是非递减的右连续函数，那么 μ 在 $\mathcal{C}$ 上是可数可加的.

87

证明　首先证明 μ 是有限可加的. 假设 $(a, b] = \bigcup_{1 \leqslant i \leqslant n} (a_i, b_i]$，这里 $(a_i, b_i]$ 是互不相交的，并假设这些区间是非空的. 不改变 $\mu((a_i, b_i])$ 的和，区间可以被重新标记，使得对 $j = 1, 2, \cdots, n$，有 $b_{j-1} = a_j$. 那么

$$G(b) - G(a) = \sum_{1 \leqslant j \leqslant n} G(b_j) - G(a_j)$$

(利用一个压缩和). 因此, μ 在$\mathcal{C}$上是有限可加的.

第二步证明 μ 是有限次可加的, 即如果 $a<b$, 并且对每个 j, $(a, b] \subset \bigcup_{i\leqslant j\leqslant n}(c_j, d_j]$, 其中 $c_j<d_j$, 那么 $G(b)-G(a) \leqslant \sum_{i\leqslant j\leqslant n} G(d_j)-G(c_j)$. 这可通过对 n 用归纳法来证. 对 $n=1$, 显然成立. 一般地, 对某个 j, 有 $c_j<b\leqslant d_j$(如果 $a=b$ 也不影响证明). 通过重新编号, 取 $j=n$. 如果$c_n\leqslant a$, 结论成立. 如果 $c_n>a$, 那么$(a, c_n]$被剩余的 $n-1$ 个区间覆盖, 因此, 由归纳假设,

$$G(b)-G(a)=G(b)-G(c_n)+G(c_n)-G(a)$$
$$\leqslant G(b)-G(c_n)+\sum_{j=1}^{n-1} G(d_j)-G(c_j).$$

下面证可数可加性. 假设区间 $J:=(c, d]$是可数多个互不相交的区间 $J_i:=(c_i, d_i]$的并. 对每个有限的 n, J 是 $J_i(i=1, 2, \cdots, n)$和有限多个与 J_i 互不相交的左开右闭区间的并. 特别地, 把区间 $(c_i, d_i](i=1, 2, \cdots, n)$重新记作$(a_i, b_i](i=1, 2, \cdots, n)$, 这里 $c\leqslant a_1<b_1\leqslant a_2<\cdots\leqslant a_n<b_n\leqslant d$. 那么其他的区间为$(c, a_1]$, $(b_j, a_{j+1}](j=1, 2, \cdots, n-1)$和$(b_n, d]$.

根据有限可加性, 对所有的 n, 有

$$\mu(J) \geqslant \sum_{i=1}^{n}\mu(J_i), \quad 因此 \quad \mu(J) \geqslant \sum_{i=1}^{\infty}\mu(J_i).$$

为证相反的不等式, 令 $J:=(c, d]\in\mathcal{C}$. 令 $\varepsilon>0$, 对每个 n, 由 G 的右连续性知, 存在 $\delta_n>0$, 使得

$$G(d_n+\delta_n) < G(d_n)+\varepsilon/2^n,$$

而且存在 $\delta>0$, 使得 $G(c+\delta)\leqslant G(c)+\varepsilon$. 现在紧闭区间$[c+\delta, d]$包含在可数多个开区间 $I_n:=(c_n, d_n+\delta_n)$的并中. 从而存在一个有限子覆盖. 因此, 由有限次可加性知,

$$G(d)-G(c)-\varepsilon \leqslant G(d)-G(c+\delta) \leqslant \sum_n G(d_n)-G(c_n)+\varepsilon/2^n,$$

88

且 $\mu(J)\leqslant 2\varepsilon+\sum_n \mu(J_n)$. 令 $\varepsilon\downarrow 0$ 完成证明. □

现在回到一般的情形, 我们有如下的扩张性质. 关于此性质的一个重要例子就是$\mathcal{A}$是由$\mathbb{R}$ 中左开右闭区间的所有有限并生成的环. 由$\mathcal{A}$生成的 σ-代数包含一些非常复杂的集合.

3.1.4 定理 对任意集 X 和 X 的子集生成的环$\mathcal{A}$, 由$\mathcal{A}$映射到$[0, \infty)$的任意可数可加函数μ 都可以扩张为由$\mathcal{A}$生成的 σ-代数$\mathcal{S}$上的一个测度.

证明 对任意集 $E\subset X$, 令

$$\mu^*(E) := \inf\{\sum_{1\leqslant n<\infty}\mu(A_n): A_n\in\mathcal{A}, E\subset\bigcup_n A_n\},$$

或 $+\infty$, 如果这样的 A_n 不存在.

那么μ^*称为由 μ 定义的外测度(outer measure). 注意到$\mu^*(\varnothing)=0$(对所有的 n, 令 $A_n=\varnothing$). 这个证明包括四个引理. 第一个引理给出了 μ^* 的可数次可加性.

3.1.5 引理 对任意集 E 和 $E_n\subset X$, 如果 $E\subset\bigcup_n E_n$, 那么 $\mu^*(E)\leqslant\sum_n\mu^*(E_n)$.

证明 如果右边的和为 $+\infty$, 显然成立. 否则, 给定 $\varepsilon>0$, 对每个 n, 取 $A_{nm}\in\mathcal{A}$, 使得 $E_n\subset\bigcup_m A_{nm}$且 $\sum_m\mu(A_{nm})<\mu^*(E_n)+\varepsilon/2^n$. 那么 E 属于 A_{mj}关于 m 和 j 的并, 因此(由引理 3.1.2),

$$\mu^*(E) \leqslant \sum_{1\leqslant n<\infty}\mu^*(E_n)+\varepsilon/2^n \leqslant \varepsilon+\sum_{1\leqslant n<\infty}\mu^*(E_n).$$

令 $\varepsilon\downarrow 0$，引理 3.1.5 成立．□

3.1.6 引理　对任意 $A\in\mathcal{A}$，$\mu^*(A)=\mu(A)$.

证明　如果 $A\subset\bigcup_n A_n$，$A_n\in\mathcal{A}$，令 $B_n:=A\cap A_n\setminus\bigcup_{j<n}A_j$，则 B_n 互不相交，在$\mathcal{A}$中，且其并为 A. 由假设，$\mu(A)=\sum_n\mu(B_n)\leqslant\sum_n\mu(A_n)$. 因此 $\mu(A)\leqslant\mu^*(A)$. 反之，取 $A_1=A$，对 $n>1$，取 $A_n=\varnothing$，证明$\mu^*(A)\leqslant\mu(A)$，于是 $\mu^*(A)=\mu(A)$. □

集合 $F\subset X$ 称为 μ^*-可测的(μ^*-measurable)，$F\in M(\mu^*)$当且仅当对每个集合 $E\subset X$，有 $\mu^*(E)=\mu^*(E\cap F)+\mu^*(E\setminus F)$. 换句话说，$F$ 可分离所有对 μ^* 可加的集合．这等价于当 $\mu^*(E)<+\infty$ 时，

$$\mu^*(E)\geqslant\mu^*(E\cap F)+\mu^*(E\setminus F),$$

由于相反的不等式总是成立(由引理 3.1.5)，从而完成证明．

89

3.1.7 引理　$\mathcal{A}\subset\mathcal{M}(\mu^*)$($\mathcal{A}$中的所有集合都是 μ^* 可测的).

证明　设 $A\in\mathcal{A}$，$E\subset X$ 满足 $\mu^*(E)<\infty$. 给定 $\varepsilon>0$，取 $A_n\in\mathcal{A}$，使得 $E\subset\bigcup_n A_n$，且 $\sum_n\mu(A_n)\leqslant\mu^*(E)+\varepsilon$. 于是有 $E\cap A\subset\bigcup_n(A\cap A_n)$，$E\setminus A\subset\bigcup_n(A_n\setminus A)$，所以

$$\mu^*(E\cap A)+\mu^*(E\setminus A)\leqslant\sum_n\mu(A\cap A_n)+\mu(A_n\setminus A)=\sum_n\mu(A_n).$$

令 $\varepsilon\downarrow 0$，有 $\mu^*(E\cap A)+\mu^*(E\setminus A)\leqslant\mu^*(E)$. □

3.1.8 引理　$\mathcal{M}(\mu^*)$是一个 σ 代数，μ^*是其上的一个测度．

证明　显然 $F\in\mathcal{M}(\mu^*)$当且仅当 $X\setminus F\in\mathcal{M}(\mu^*)$. 如果 A，$B\in\mathcal{M}(\mu^*)$，那么对任意 $E\subset X$，有$A\cup B=X\setminus[(X\setminus A)\cap(X\setminus B)]$，而且

$$\begin{aligned}\mu^*(E)&=\mu^*(E\cap A)+\mu^*(E\setminus A) &&\text{因为 } A\in\mathcal{M}(\mu^*)\\&=\mu^*(E\cap A\cap B)+\mu^*(E\cap A\setminus B)+\mu^*(E\setminus A) &&\text{因为 } B\in\mathcal{M}(\mu^*)\\&=\mu^*(E\cap(A\cap B))+\mu^*(E\setminus(A\cap B)) &&\text{因为 } A\in\mathcal{M}(\mu^*).\end{aligned}$$

因此$\mathcal{M}(\mu^*)$是一个代数．设 $E_n\in\mathcal{M}(\mu^*)$，$n=1,2,\cdots$，$F:=U_{1\leqslant j<\infty}E_j$，且 $F_n:=\bigcup_{1\leqslant j\leqslant n}E_j\in\mathcal{M}(\mu^*)$. 由于对所有的 n，$E_n\setminus\bigcup_{j<n}E_j\in\mathcal{M}(\mu^*)$，所以在证明 F 可测时，假设 E_n 互不相交．对任意 $E\subset X$，有

$$\begin{aligned}\mu^*(E)&=\mu^*(E\setminus F_n)+\mu^*(E\cap F_n) &&\text{因为 } F_n\in\mathcal{M}(\mu^*)\\&=\mu^*(E\setminus F_n)+\mu^*(E\cap E_n)+\mu^*(E\cap\bigcup_{j<n}E_j) &&\text{因为 } E_n\in\mathcal{M}(\mu^*).\end{aligned}$$

因此，对 n 作归纳，

$$\mu^*(E)=\mu^*(E\setminus F_n)+\sum_{j=1}^n\mu^*(E\cap E_j)\geqslant\mu^*(E\setminus F)+\sum_{j=1}^n\mu^*(E\cap E_j).$$

当 $n\to\infty$ 时，

$$\mu^*(E)\geqslant\mu^*(E\setminus F)+\sum_{j=1}^\infty\mu^*(E\cap E_j)\geqslant\mu^*(E\setminus F)+\mu^*(E\cap F).$$

根据引理 3.1.5，$F\in\mathcal{M}(\mu^*)$，再次利用引理 3.1.5 得，

$$\mu^*(E)=\mu^*(E\setminus F)+\sum_{j\geqslant 1}\mu^*(E\cap E_j).$$

令 $E=F$，则可证得 μ^* 在$\mathcal{M}(\mu^*)$上是可数可加的，由此就证明了引理 3.1.8 和定理 3.1.4. □

90

外测度为 0 的集合总是可测的.

3.1.9 命题 如果 $\mu^*(E)=0$，那么 $E\in\mathcal{M}(\mu^*)$.

证明 对任意 $A\in X$,

$$\mu^*(A)\geqslant\mu^*(A\setminus E)=\mu^*(A\setminus E)+\mu^*(E)\geqslant\mu^*(A\setminus E)+\mu^*(A\cap E).$$ □

集合 A 和 B 的对称差(symmetric difference)定义为 $A\Delta B:=(A\setminus B)\cup(B\setminus A)$. 因而，$A\Delta B$ 是 A 或 B 中的点，但不同时是 A 和 B 中的点. 例如，$[0,2]\Delta[1,3]=[0,1)\cup(2,3]$.

X 的子集构成的集类$\mathcal{A}$上的函数 μ 称为是 σ-有限的(σ-finite)，当且仅当对所有的 n 和 $X=\bigcup_n A_n$，有$\{A_n\}\in\mathcal{A}$，$|\mu(A_n)|<\infty$. 例如，如果$\mathcal{A}$是实数集$\mathbb{R}$上满足 $\lambda((a,b]):=b-a$ 的所有区间$(a,b]$构成的集类，其中 $a<b$，那么 λ 是 σ-有限的. 定义在由集合 X 的所有子集构成的集类上的计数测度是 σ-有限的，当且仅当 X 是可数的. 或者对某个集类中的所有非空子集 A，如果 $\mu(A)=+\infty$ 且 $\mu(\varnothing)=0$，则 μ 显然不是 σ-有限的. 另一个例子是，令$\mathcal{A}$是由所有区间$(a,b]$生成的代数，对$\mathcal{A}$中的所有非空集合，令$\mu(A)=+\infty$. 设$\mathcal{B}$是由$\mathcal{A}$生成的 σ 代数，则定义在$\mathcal{B}$上且在$\mathcal{A}$上测度等于 μ 的测度是计数测度. 另外一个例子是，将测度 $+\infty$ 分配到$\mathcal{B}$上的所有非空集合中. 下面的唯一性定理说明了 σ-有限性是多么有用.

3.1.10 定理 设 μ 是代数$\mathcal{A}$上的非负可数可加测度，α 是由$\mathcal{A}$生成的 σ-代数$\mathcal{S}$上的一个测度，满足在$\mathcal{A}$上，$\alpha=\mu$，则对任意满足 $\mu^*(A)<+\infty$ 的 $A\in\mathcal{S}$，有 $\alpha(A)=\mu^*(A)$. 如果 μ 是 σ-有限的，那么从代数$\mathcal{A}$到其生成的 σ-代数$\mathcal{S}$(由定理 3.1.4 给出)，μ 的扩张测度是唯一的，且在$\mathcal{S}$上扩张测度 α 与 μ^* 相等.

证明 对任意满足 $A\subset\bigcup_n A_n$ 的 $A\in\mathcal{S}$和 $A_n\in\mathcal{A}$，由引理 3.1.6 的证明知，$\alpha(A)\leqslant\sum_n\alpha(A_n)=\sum_n\mu(A_n)$，取下确界，有 $\alpha(A)\leqslant\mu^*(A)$.

如果 $\mu^*(A)<\infty$，那么给定 $\varepsilon>0$，取 A_n 使得 $\sum_n\mu(A_n)<\mu^*(A)+\varepsilon/2$. 令 $B_k:=\bigcup_{1\leqslant m<k}A_m$，则对 $k<\infty$，有 $B_k\in\mathcal{A}$，且 $B_\infty\in\mathcal{S}$. 由于 $A\subset B_\infty$，

$$\mu^*(A)\leqslant\mu^*(B_\infty)<\mu^*(A)+\varepsilon/2.$$

由引理 3.1.7 和引理 3.1.8 知，这意味着$\mathcal{S}\subset\mathcal{M}(\mu^*)$，由定理 3.1.1 知，对充分大的 k，$\mu^*(B_\infty\setminus B_k)<\frac{\varepsilon}{2}$ 且 $\mu^*(B_\infty\setminus A)=\mu^*(B_\infty)-\mu^*(A)<\varepsilon/2$. 从而 $\alpha(B_\infty)\leqslant\mu^*(B_\infty)<\infty$ 因为在$\mathcal{S}$上，$\alpha\leqslant\mu^*$，所以

91

$$\begin{aligned}\alpha(A)&=\alpha(B_\infty)-\alpha(B_\infty\setminus A)\geqslant\alpha(B_k)-\mu^*(B_\infty\setminus A)\\&\geqslant\alpha(B_k)-\varepsilon/2=\mu^*(B_k)-\varepsilon/2\quad(\text{因为 }B_k\in\mathcal{A}\text{以及引理 3.1.6})\\&\geqslant\mu^*(B_\infty)-\varepsilon\geqslant\mu^*(A)-\varepsilon.\end{aligned}$$

令 $\varepsilon\downarrow0$，可得 $\alpha(A)\geqslant\mu^*(A)$，从而 $\alpha(A)=\mu^*(A)$. 或者，如果 A_n 是互不相交的且 $\mu(A_n)<\infty$，那么由引理 3.1.8，

$$\alpha(A)=\sum_n\alpha(A\cap A_n)=\sum_n\mu^*(A\cap A_n)=\mu^*(A).$$

因此，如果 μ 是 σ-有限的，则 α 是唯一的. □

下面证明，即使在可能不可测的集合上，外测度 μ^* 也有下连续的性质.

***3.1.11 定理** 设 μ 是由集合 X 的子集构成的代数$\mathcal{A}$上的非负可数可加测度，μ^* 是通过 μ 定义的外测度，设 A_n 和 A 是 X 的任意子集(不一定可测)，使得当 $n\to\infty$ 时，$A_n\uparrow A$，这意味着 $A_1\subset A_2\subset\cdots$ 且

$\bigcup_n A_n = A$，那么 $\mu^*(A_n) \uparrow \mu^*(A)$.

例：设 $B_1, \cdots, B_k$ 均为非空互不相交的子集，且其并为 X. 设$\mathcal{A}$是由$\{B_1, \cdots, B_k\}$生成的代数，那么$\mathcal{A}$是一个 σ-代数，且存在$\mathcal{A}$上的测度 μ，使得 $\mu(B_j) = 1$，$j = 1, \cdots, k$. 对任意 $A \subset X$，$\mu^*(A)$是 B_j 的数量，其中 A 是交集，换句话说，$\mu^*(A)$是满足 $A \cap B_j \neq \varnothing$ 的 j 的取值个数．在定理 3.1.11 的条件下，对充分大的 n，$\mu^*(A_n) = \mu^*(A)$.

证明 由 μ^* 的定义，当 $B \subset C$ 时，$\mu^*(B) \leqslant \mu^*(C)$，所以 $\mu^*(A_n)$关于 n 是非递减的，并达到某一极限 $c \leqslant \mu^*(A)$. 因此，我们需要证明相反的不等式成立．如果当 $n \to \infty$ 时，有 $\mu^*(A_n) \to +\infty$，结论显然成立．因此，我们可假设 $c < \infty$. 给定 $\varepsilon > 0$，由 μ^* 的定义，存在集合序列 $A_{nj} \in \mathcal{A}$，满足对所有的 j，$A_n \subset B_n := \bigcup_j A_{nj}$ 且 $\sum_j \mu(A_{nj}) < \mu^*(A_n) + \varepsilon$. 如果 $B_{nk} := \bigcup_{1 \leqslant j < k} A_{nj}$ 且由定理 3.1.4，将 μ 在由$\mathcal{A}$生成的 σ-代数$\mathcal{L}$上的扩张测度仍记为 μ，我们有

$$\mu(B_n) = \mu(A_{n1}) + \sum_{k \geqslant 2} \mu(A_{nk} \setminus B_{nk}) \leqslant \sum_j \mu(A_{nj}) < \mu^*(A_n) + \varepsilon.$$

92

令 $C_n := \bigcap_{r \geqslant n} B_r$，则 $A_n \subset C_n$，$C_n \in \mathcal{L}$，$\mu^*(C_n) < \mu^*(A_n) + \varepsilon$ 且 $C_1 \subset C_2 \subset \cdots$. 令 $C := \bigcup_n C_n$，则由定理 3.1.1，$\mu(C_n) \uparrow \mu(C) \leqslant c + \varepsilon$. 又因为 $A \subset C$，所以 $\mu^*(A) \leqslant \mu(C) \leqslant c + \varepsilon$. 令 $\varepsilon \downarrow 0$ 有 $\mu^*(A) \leqslant c$，从而 $\mu^*(A) = c$. □

习题

1. 设$(X, \mathcal{S}, \mu)$是 σ-有限测度空间，满足 $\mu(X) = +\infty$，证明：对任意 $M < \infty$，存在某个 $A \in \mathcal{S}$，使得 $M < \mu(A) < \infty$.
2. 设 X 是无限集，对任意有限集 A，有 $m(A) = 0$，对任意无限集 A，有 $m(A) = +\infty$，证明：m 是有限可加但不是可数可加的．
3. 设 X 是无限集，$\mathcal{A}$是由 X 的子集 A 组成的集类，使得若 A 有限，则 $m(A) := 0$，若 A 的补集有限，则 $m(A) := 1$.
   (a) 证明：$\mathcal{A}$是一个代数但不是 σ-代数．
   (b) 证明：m 在$\mathcal{A}$上是有限可加的．
   (c) 在 X 满足什么条件下，m 可以扩张为 σ-代数上的可数可加测度？
4. 设$(X, \mathcal{T})$是一个拓扑空间，$p \in X$. p 的一个击穿邻域为集合 B，对某个 p 的邻域 A，B 包括 $A \setminus \{p\}$. 换句话说，B 是 p 的邻域，除了 B 不可能包括 p 本身外．如果 B 是 p 的击穿邻域，令 $m(B) = 1$，如果 B 与 p 的某个击穿邻域不相交，令 $m(B) = 0$.
   (a) 证明：m 定义在一个代数上．
   (b) 如果(S, d)是度量空间，且$\{p\}$是非开的，证明：m 是有限可加的．
   (c) 如果(S, d)是度量空间，证明：不能把 m 扩张成一个测度．
5. 在 $I := [0, 1]$ 上，设$\mathcal{F}$是所有第一范畴集构成的集类（无处稠密集的可数并，参见 2.5 节）．对每个 $A \in \mathcal{F}$，令 $m(A) = 0$，$m(I \setminus A) = 1$.
   (a) 证明：m 是定义在 σ-代数上的测度．
   (b) 设 $B = [0, 1/2]$，问 B 是 m^* 可测的吗？
6. (a) 设$\mathcal{S}$是 X 的子集的 σ-代数，$\mu_1, \cdots, \mu_n$ 是$\mathcal{S}$上的测度，令 $c_j (j = 1, \cdots, n)$ 为非负常数，证明：$c_1\mu_1 + \cdots + c_n\mu_n$ 是$\mathcal{S}$上的测度．
   (b) 对任意 $x \in X$ 及 X 的子集 A，设 $\delta_x(A) := 1_A(x)$（若 $x \in A$，则 $\delta_x(A) = 1$，否则为 0），证明：δ_x 是 2^X 上的

93 测度．

7. 给定测度空间$(X, \mathcal{S}, \mu)$，设A_j，B_j是X的子集，满足$\mu^*(A_j \Delta B_j)=0$，$j=1, 2, \cdots$，证明：$\mu^*\left(\bigcup_j A_j\right)=\mu^*\left(\bigcup_j B_j\right)$.
8. 设$\mathcal{A}$是$\mathbb{R}$的子集的代数，其中$\mathbb{R}$是由满足$A_{a,b}:=[-b, -a)\cup(a, b](0<a<b)$的所有集合构成的集类．令$m(A_{a,b}):=b-a$. 证明：$m$在$\sigma$-代数上可以扩张为一个可数可加的测度．问$[1, 2]$是$m^*$可测的吗？
9. 对任意的$A\subset\mathbb{N}$，设$\#(A, n)$是A中比n小的元素的个数，且对任意的$A\in\mathcal{C}$，$d(A):=\lim\limits_{n\to\infty}\#(n, A)/n$,其中对于集类$\mathcal{C}$，极限存在．(这里$d(A)$称为$A$的密度(density).)

   (a)证明：d在$\mathcal{C}$上不是可数可加的.

   (b)证明：$\mathcal{C}$不是一个环．[提示：令A是偶数的集合，B是2^{2n}和2^{2n+1}之间的偶数与2^{2n-1}和2^{2n}之间的奇数的集合，其中$n=1, 2, \cdots$，A，B和$A\cap B$哪个在$\mathcal{C}$中？(顺便提一下，d在$2^{\mathbb{N}}$上可扩张成有限可加的．)]
10. (a)设$\mathcal{U}$是集合X的子集的一个超滤子．对所有的$A\in\mathcal{U}$，令$m(A)=1$，否则，$m(A)=0$. 证明：m在2^X上是有限可加的．

    (b)设X是一个无限集，m在由2^X到$\{0, 1\}$上是有限可加的且$m(X)=1$. 证明：$\{A: m(A)=1\}$是一个超滤子．

    (c)证明：对任意无限集，存在一个定义在2^X上的有限可加函数m，满足对任意有限集A，$m(A)=0$且$m(X)=1$.[提示：利用(a).]

    (d)如果X是可数的，证明：(c)中的m不是可数可加的．

    (e)在一个游戏中，Sam 和 Joe 分别随机挑选一个非负整数．每人挑选任意集合A中的数的概率为$m(A)$，m为(c)中所定义，$X=\mathbb{N}$．选出较大数者胜出．抛硬币决定你最先看见谁的数．游戏进行中，你看见 Sam 的数但始终不知 Joe 的数．那么你认为谁会胜出？(在无可数可加性时，怪事发生．)

3.2 半环和环

在3.1节中，环上的可数可加非负函数在定理3.1.4中被扩张成σ-代数上的某个函数(测度). 在定理3.1.3中，证明了在由所有左开右闭区间$(a, b]:=\{x\in\mathbb{R}: a<x\leqslant b\}$构成的集类$\mathcal{C}$上，长度是可数可加的．由于$\mathcal{C}$不是环，使得这些事实之间缺少联系．为寻找其间的联系，接下来证明$\mathcal{C}$具有如下性质，这在扩张测度中更为有用． 94

定义 对任意集合X，集类$\mathcal{D}\subset 2^X$称为**半环**(semiring)，如果$\varnothing\in\mathcal{D}$且对任意的$A$，$B\in\mathcal{D}$，有$A\cap B\in\mathcal{D}$，对某些有限的$n$和互不相交的$C_j\in\mathcal{D}$，有$A\setminus B=\bigcup_{1\leqslant j\leqslant n}C_j$.

3.2.1 命题 $\mathcal{C}$是半环．

证明 我们有$\varnothing=(a, a]\in\mathcal{C}$，$\mathcal{C}$中两个区间的交仍在$\mathcal{C}$中(可能为空集). 现在$(a, b]\setminus(c, d]$或者在$\mathcal{C}$中，或者如果$a<c<d<b$，则它是$\mathcal{C}$中两个互不相交的集合的并． □

另一种半环在关于笛卡儿积构造测度中是很有用的．如果假设$\mathcal{A}$和$\mathcal{B}$是区间的半环，那么在下面的事实中$\mathcal{D}$将是矩形的半环．

3.2.2 命题 设$X=Y\times Z$，$\mathcal{A}$和$\mathcal{B}$分别是集合Y和Z的子集构成的半环，令$\mathcal{D}:=\{A\times B: A\in\mathcal{A}, B\in\mathcal{B}\}$，那么$\mathcal{D}$是半环．

证明 首先$\varnothing=\varnothing\times\varnothing\in\mathcal{D}$. 对"交"运算：对任意的$A$，$E\in\mathcal{A}$和$B$，$F\in\mathcal{B}$，有

$$(A\times B)\cap(E\times F)=(A\cap E)\times(B\cap F).$$

对"差"运算：

$$D := (A\times B)\setminus(E\times F) = (A\times B)\setminus((A\cap E)\times(B\cap F)),$$

所以在求 D 的值时，我们可假设 $E\subset A$，$F\subset B$，于是对一些有限的 m，n，互不相交的 $G_j\in\mathcal{A}$及互不相交的 $H_k\in\mathcal{B}$，有

$$D = ((A\setminus E)\times B)\cup(E\times(B\setminus F)) = \Big(\Big(\bigcup_{1\le j\le m} G_j\Big)\times B\Big)\cup\Big(E\times\bigcup_{1\le k\le n} H_k\Big)$$

$$= \bigcup_{1\le j\le m}(G_j\times B)\cup\bigcup_{1\le k\le n}(E\times H_k).$$

因此 D 是$\mathcal{D}$中互不相交集合的并(因为对所有的 j，$G_j\cap E=\varnothing$)．□

95

注意到，称半环$\mathcal{R}$为环，如果对任意 A，$B\in\mathcal{R}$，有 $A\cup B\in\mathcal{R}$. 任一半环给出如下的环．

3.2.3 命题　对任意半环$\mathcal{D}$，令$\mathcal{R}$是$\mathcal{D}$中元素的所有有限互不相交的并构成的集合，则$\mathcal{R}$是环．

证明　显然$\varnothing\in\mathcal{R}$. 令 $A=\bigcup_{1\le j\le m}A_j$，$B=\bigcup_{1\le k\le n}B_k$，其中 A_j，B_k 均为$\mathcal{D}$中互不相交的集合，那么

$$A\cap B = \cup\{A_j\cap B_k : j=1,2,\cdots,m, k=1,2,\cdots,n\},$$

其中 $A_j\cap B_k$ 在$\mathcal{D}$中且互不相交，所以 $A\cap B\in\mathcal{R}$. $\mathcal{R}$中有限多个集合的不相交并显然在$\mathcal{R}$中．因为 $A\cup B=A\cup(B\setminus A)$，所以接下来只需证明 $B\setminus A\in\mathcal{R}$. 由于

$$B\setminus A = \bigcup_{1\le k\le n}\Big(B_k\setminus\bigcup_{1\le j\le m}A_j\Big),$$

是关于 k 的不相交并，所以这足以证明对每个 k，$B_k\setminus\bigcup_{1\le j\le m}A_j=\bigcap_{1\le j\le m}B_k\setminus A_j\in\mathcal{R}$. 由半环的定义，每个 $B_k\setminus A_j\in\mathcal{R}$. $\mathcal{R}$中任两个集合的交还在$\mathcal{R}$中，由归纳法，$\mathcal{R}$中有限多个集合的交也在$\mathcal{R}$中．□

例如，任意两个区间$(a, b]$和$(c, d]$的并，如果两区间相交，其并是另外一个左开右闭区间$(\min(a, c), \max(b, d)]$，如果 $b=c$ 或者 $a=d$，其并是一个不交并，或者两者都可能．

对任意集类$\mathcal{A}\in 2^X$，包含$\mathcal{A}$的所有环的交仍是一个包含$\mathcal{A}$的环，称之为由$\mathcal{A}$生成的环．如果$\mathcal{A}$是一个半环，显然，由$\mathcal{A}$生成的环就是命题 3.2.3 中给出的环．由此，从半环到环的可加函数产生了．

3.2.4 命题　设$\mathcal{D}$是任意的半环，α 是从$\mathcal{D}$映射到$[0,\infty]$的可加函数．对互不相交的 $C_j\in\mathcal{D}$，令

$$\mu\Big(\bigcup_{1\le j\le m}C_j\Big) := \sum_{1\le j\le m}\alpha(C_j).$$

那么 μ 在由$\mathcal{D}$生成的环$\mathcal{R}$上是有定义且可加的．如果 α 在$\mathcal{D}$上是可数可加的，则 μ 在$\mathcal{R}$上也是有定义且可加的，从而 μ 可以扩张为由$\mathcal{D}$或$\mathcal{R}$生成的 σ-代数上的测度．

证明　假设 $C_j\in\mathcal{D}$是互不相交的，某些 D_k 也是互不相交的，且 $\bigcup_{1\le j\le m}C_j=\bigcup_{1\le k\le n}D_k$. 那么对每个 j，$C_j=\bigcup_{1\le k\le n}C_j\cap D_k$ 为$\mathcal{D}$中集合的不交并，所以由可加性，

96

$$\sum_{j=1}^{m}\alpha(C_j) = \sum_{j=1}^{m}\sum_{k=1}^{n}\alpha(C_j\cap D_k) = \sum_{k=1}^{n}\alpha(D_k).$$

因此，μ 是明确定义的．对$\mathcal{R}$中任意互不相交的 B_1，$\cdots$，B_n，令 $B_i=\bigcup_{1\le j\le m(i)}B_{ij}$，$i=1, 2, \cdots, n$，其中对每个 i. $B_{ij}\in\mathcal{D}$不相交，且 $m(i)<\infty$. 从而所有的 B_{ij}互不相交．令 B_{ij}关于所有的 i 和 j 求和，所得和为 B. 因为 μ 是明确定义的，

$$\mu(B) = \sum_{i=1}^{n}\sum_{j=1}^{m(i)}\alpha(B_{ij}) = \sum_{i=1}^{n}\mu(B_i),$$

因此 μ 是可加的．假设 α 是可数可加的，$B\in\mathcal{R}$，因此，对某些不相交的 $C_i\in\mathcal{D}$，有 $B=\bigcup_{1\le i\le n}C_i$ 且对

某些不相交的 $A_r\in\mathcal{R}$，有 $B=\bigcup_{1\le r<\infty}A_r$. 对某些不相交的 $A_{rj}\in\mathcal{D}$，令 $A_r=\bigcup_{1\le j\le k(r)}A_{rj}$，则

$$B=\bigcup_{i=1}^{n}\bigcup_{r=1}^{\infty}\bigcup_{j=1}^{k(r)}C_i\cap A_{rj}$$

为$\mathcal{D}$中不交集的并．每个 A_r 是集合 $C_i\cap A_{rj}$ 的一个有限不交并．由 α 在$\mathcal{D}$上的可数可加性和引理 3.1.2 重排非负项的和得，

$$\mu(B)=\sum_{i=1}^{n}\alpha(C_i)=\sum_{i=1}^{n}\sum_{r=1}^{\infty}\sum_{j=1}^{k(r)}\alpha(C_i\cap A_{rj})$$
$$=\sum_{r=1}^{\infty}\sum_{i=1}^{n}\sum_{j=1}^{k(r)}\alpha(C_i\cap A_{rj})=\sum_{r=1}^{\infty}\mu(A_r),$$

因此，μ 在$\mathcal{R}$上是可数可加的．从而由定理 3.1.4 知，μ 可以扩张成 σ-代数上的测度． □

代数$\mathcal{A}$是满足$\mathcal{A}$中任意集合的补集仍在$\mathcal{A}$中的环．为得到一个包含环$\mathcal{R}$的最小代数，我们必须把$\mathcal{R}$中集合的补集也放到$\mathcal{R}$中．例如，由于补集的并是交的补集，从而使得这种做法是合理的.

3.2.5 命题 *设$\mathcal{R}$是由集合 X 的子集构成的任意环，令$\mathcal{A}:=\mathcal{R}\cup\{X\setminus B: B\in\mathcal{R}\}$，那么$\mathcal{A}$是一个代数．*

97

证明 显然$\varnothing\in\mathcal{A}$，且 $A\in\mathcal{A}$当且仅当 $X\setminus A\in\mathcal{A}$. 令 $C\in\mathcal{A}$，$D\in\mathcal{A}$，则 $C\cap D\in\mathcal{A}$，如果

(a)C，$D\in\mathcal{R}$，有 $C\cap D\in\mathcal{R}$.

(b)$C\in\mathcal{R}$且 $D=X\setminus B$，$B\in\mathcal{R}$，满足 $C\cap D=C\setminus B\in\mathcal{R}$.

(c)$X\setminus C$，$X\setminus D\in\mathcal{R}$，此时 $X\setminus(C\cap D)=(X\setminus C)\cup(X\setminus D)\in\mathcal{R}$. □

正如 3.1 节中那样，现在令 G 是从$\mathbb{R}$映射到自身的非递减右连续函数．最重要的特殊情形是，当 G 是恒等函数时，$G(x)\equiv x$. 为简单起见，在这种情形下考虑可能更容易．在所有区间$(a,b]$的半环$\mathcal{C}$上，对 $a\le b$，a，$b\in\mathbb{R}$，由 $\mu((a,b])=G(b)-G(a)$知，$\mu:=\mu_G$. 如果 G 是恒等函数，那么称 μ 为勒贝格(长度)λ.

从命题 3.2.3 知，由$\mathcal{C}$生成的环包含所有有限并

$$A=\bigcup_{j=1}^{n}(a_j,b_j],\quad -\infty<a_1\le b_1<\cdots\le b_n<+\infty.$$

A 的补集为

$$(-\infty,a_1]\cup(b_1,a_2]\cup\cdots\cup(b_n,+\infty).$$

令$\mathcal{B}$为由$\mathcal{C}$生成的$\mathbb{R}$的子集的 σ-代数，注意，任意开区间$(a,b)(b<\infty)$为$\mathcal{B}$中区间$(a,b-1/n]$$(n=1,2,\cdots)$的并．对有理数 q，r，开区间(q,r)形成了$\mathbb{R}$的拓扑的一个可数基，因此，所有开集都在$\mathcal{B}$中．反之，任意区间$(a,b]$是开区间$(a,b+1/n](n=1,2,\cdots)$的交，则$\mathcal{B}$也是由$\mathbb{R}$的拓扑生成的 σ-代数．在任意的拓扑空间中，由拓扑生成的 σ-代数称为博雷尔 σ-代数，代数中的集合称为博雷尔集．

下面的定理说明了对每个函数 G 产生的一个测度．

3.2.6 定理 *对任意的非递减右连续函数 G，在区间$(a,b](a\le b)$构成的集类$\mathcal{C}$上，$\mu_G((a,b]):=G(b)-G(a)(a\le b)$能唯一地扩张成博雷尔 σ-代数上的测度．*

特别地，勒贝格测度 λ 存在，也就是对任意实数 $a\le b$，$\mathbb{R}$中的博雷尔 σ-代数上的测度满足 $\lambda((a,b])=b-a$.

证明 首先由定理 3.1.3 知，μ_G 在$\mathcal{C}$上是可数可加的．由命题 3.2.1 知，$\mathcal{C}$是一个半环．因此根据命题 3.2.4，μ_G 可扩张为博雷尔 σ-代数上的一个测度．因为对所有的 $n=0,\pm1,\pm2,\cdots$，

$\mu_G((n, n+1]) < \infty$，所以 μ_G 是 σ-有限的，由定理 3.1.10 知，在每个$(n, n+1]$上，μ_G 具有唯一性且是可数可加的．□

98

更一般地，什么情况下，σ-代数$\mathcal{B}$上的测度可由其在一个属于类$\mathcal{C} \subset \mathcal{B}$的集合上的值唯一确定？如果对每个在所有属于$\mathcal{C}$的集合上具有有限值的测度$\mu$，这都成立，那么称$\mathcal{C}$为*确定类*(determining class)．例如，如果 X 是有限集且$\mathcal{B} = 2^X$ 为其子集构成的集类，那么由所有单元素集$\{x\}$ $(x \in X)$构成的集类$\mathcal{C}$是一个确定类．下面是使类$\mathcal{C}$成为确定类的两个充分条件．

3.2.7 定理　*如果 X 是一个集合，$\mathcal{C} \subset \mathcal{B}$，其中$\mathcal{B}$为由$\mathcal{C}$生成的 X 的子集的 σ-代数(即包括$\mathcal{C}$的最小 σ-代数)，如果$\mathcal{C}$是一个代数，或者 X 是$\mathcal{C}$中可数多个集合的并且$\mathcal{C}$为半环，则$\mathcal{C}$为一个确定类．*

证明　如果μ 为$\mathcal{B}$上的测度，且在所有属于$\mathcal{C}$的集合上有限，则由假设知，μ 为 σ-有限的．如果$\mathcal{C}$为半环，则由命题 3.2.3 知，μ 在$\mathcal{C}$上的值唯一地决定μ 在包含$\mathcal{C}$的最小环上的值，因此可假设$\mathcal{C}$为环．由假设，存在 $A_n \in \mathcal{C}$，$A_n \uparrow X$，对任意 $B \in \mathcal{B}$，$\mu(B \cap A_n) \uparrow \mu(B)$，所以只需证明$\mu(B \cap A_n)$的值是唯一的，并且可假设$\mu(X) < \infty$．由命题 3.2.5 知，$\mu$ 是唯一的且在包括$\mathcal{C}$的最小代数上具有有限值，因此可假设$\mathcal{C}$是一个代数．那么由定理 3.1.10 即可得证．□

讨论半环、环和代数是为讨论扩张的唯一性或者确定类的性质作准备．即使$\mathcal{C}$仅仅包含两个集合，$\mathcal{C}$也不能恰好生成$\mathcal{B}$.

例：令 $X = \{1, 2, 3, 4\}$，$\mathcal{B} = 2^X$，$\mathcal{C} = \{\{1, 2\}, \{1, 3\}\}$，于是$\mathcal{C}$生成$\mathcal{B}$，但$\mathcal{C}$不是一个确定类．如果对于 $j = 1, 2, 3, 4$，$\mu\{j\} = 1/4$，$\alpha\{1\} = \alpha\{4\} = 1/6$，而 $\alpha\{2\} = \alpha\{3\} = 1/3$，则在$\mathcal{C}$上 $\alpha = \mu$.

尽管勒贝格测度的构造不是必须的，但是这种构造在其他种类的集类上有时是很有用的．集类$\mathcal{L}$称为格(lattice)，当且仅当对每个 $A \in \mathcal{L}$和 $B \in \mathcal{L}$，有 $A \cap B \in \mathcal{L}$且 $A \cup B \in \mathcal{L}$. 例如，在任意拓扑空间中，拓扑(由所有开集合组成的集类)是格；由闭集构成的集类也是格．下面的命题描述了由格生成环.

***3.2.8 命题**　*令$\mathcal{L}$是集格，且$\varnothing \in \mathcal{L}$. 包含$\mathcal{L}$的最小环$\mathcal{R}$是由所有有限并 $\bigcup_{1 \le i \le n} A_i \setminus B_i$ 构成的集类$\mathcal{C}$，其中$A_i \in \mathcal{L}$，$B_i \in \mathcal{L}$，且集合 $A_i \setminus B_i$ 是不相交的．*

99

证明　显然$\mathcal{L} \subset \mathcal{C} \subset \mathcal{R}$. 下面只需证明$\mathcal{C}$是环，已知$\varnothing \in \mathcal{C}$，令 $S_m := \bigcup_{1 \le i \le m}(A_i \setminus B_i)$，$T_n := \bigcup_{i \le j \le n}(C_j \setminus D_j)$，其中 A_i，B_i，C_i，$D_i \in \mathcal{L}$，集合 $A_i \setminus B_i$ 互不相交，集合 $C_j \setminus D_j$ 互不相交．为证明 $S_m \setminus T_n \in \mathcal{C}$，对 n 应用归纳法，注意到，对任意集合 U，V 和 W，有 $U \setminus (V \cup W) = (U \setminus V) \setminus W$，所以对 $n = 1$ 时的证明将给出同样的归纳步．由于集合$(A_i \setminus B_i) \setminus (C_1 \setminus D_1)$是不相交的，可以分开地讨论它们．对任意集合 A，B，C，D，

$$(A \setminus B) \setminus (C \setminus D) = (A \setminus (B \cup C)) \cup ((A \cap C \cap D) \setminus B)$$

为不相交集合的并，所以 $S_m \setminus T_n \in \mathcal{C}$.

又 $S_m \cup T_n = S_m \cup (T_n \setminus S_m)$为$\mathcal{C}$中不相交集合的并，所以 $S_m \cup T_n \in \mathcal{C}$且$\mathcal{C}$是环．□

习题

1. (a)证明：$\mathbb{R}$ 中所有开区间(a, b)组成的集合既不是半环也不是格.
   (b)证明：由所有闭区间$[a, b]$、开区间(a, b)、左开右闭区间$(a, b]$或左闭右开区间$[a, b)$组成的集合是半环．
2. 如果$\mathcal{D}$是$\mathbb{R}^2$ 上所有形如$(a, c] \times (b, d]$的矩形区间组成的集类，证明：$\mathcal{D}$是半环，并求出半环的定义中 n 的最小可能值．

3. 把习题 2 中的集类改为$\mathbb{R}^3$上所有形如$(a, b]\times(c, d]\times(e, f]$的集合组成集类，证明结论仍成立.
4. 称集类$\mathcal{D}$是半环集(demiring)，如果对任意A，$B\in\mathcal{D}$，$A\setminus B$(如同对半环那样)和$A\cap B$是由$\mathcal{D}$中集合的有限不相交并. 令$\mathcal{D}$是由所有集合$h(a, b):=\{\theta\in\mathbb{R}$：对一些整数$n\in\mathbb{Z}$，$a<\theta+2n\pi\leqslant b\}$组成的集类，其中$a$，$b\in\mathbb{R}$，$a\leqslant b$. 证明：$\mathcal{D}$是半环集，但不是半环.
5. 令$\mathcal{D}$是半环集(定义同习题 4)且$\mathcal{R}$是$\mathcal{D}$中元素的有限不相交并组成的集类，证明$\mathcal{R}$是环.
6. 用半环集替换半环将命题 3.2.4 推广到半环集上.
7. 设$G(x)=0(x<1)$，$G(x)=1(1\leqslant x<2)$，$G(X)=3(x\geqslant 2)$. 令μ是定理 3.2.6 中由G得到的测度，试求点式群体δ_x所构成的有界线性组合μ，其中δ_x如 3.1 节习题 6 所定义.

[100]

8. 在由$\mathbb{R}^2$中的开集组成的集类中，是否存在$\mathbb{R}^2$中拓扑的基？是否存在半环？为什么？
9. 证明：由所有半直线$(-\infty, x](x\in\mathbb{R})$组成的集类是确定类(对于半直线生成的$\sigma$-代数).

3.3　测度的完备化

设μ是集合X的子集生成的σ-代数$\mathcal{S}$上的测度，从而$(X, \mathcal{S}, \mu)$是一个测度空间. 如果$\mathcal{S}$不完全是幂集2^X，那么可把它扩张到一个较大的σ-代数上，这是有用的，特别是一些集合的测度在一定程度上可被$\mathcal{S}$中原有集合的测度确定时. 例如，如果$A\subset E\subset C$，A，$C\in\mathcal{S}$，但$E\notin\mathcal{S}$，$\mu(A)=\mu(C)$，似乎$\mu(E)$和$\mu(A)$应该相等.

本节将对μ作上述的扩张. 采用的方法与 3.1 节所定义的外测度μ^*有关. 对任意集合E，首先是找一个集合$C\in\mathcal{S}$，满足$E\subset C$且$\mu(C)$尽可能的小.

3.3.1 定理　*对任意的测度空间$(X, \mathcal{S}, \mu)$和任意的$E\subset X$，都有$C\in\mathcal{S}$且$E\subset C$，$\mu^*(E)=\mu(C)$.*

证明　如果$E\subset\bigcup_{n\geqslant 1}A_n$，$A_n\in\mathcal{S}$，令$A:=\bigcup_{n\geqslant 1}A_n$，则$E\subset A$，$A\in\mathcal{S}$，且由引理 3.1.6 的证明，$\mu(A)\leqslant\sum_{n\geqslant 1}\mu(A_n)$. 因此，$\mu^*(E)=\inf\{\mu(A): E\subset A, A\in\mathcal{S}\}$. 对每个$m=1, 2, \cdots$，选择$C_m\in\mathcal{S}$，使得$E\subset C_m$，$\mu(C_m)\leqslant\mu^*(E)+1/m$. 令$C:=\bigcap_m C_m$，则$C\in\mathcal{S}$，因此，$C=X\setminus\left(\bigcup_m X\setminus C_m\right)$，$E\subset C$，$\mu^*(E)=\mu(C)$. □

定理 3.3.1 中的集合C称为集合E的*可测覆盖*(measurable cover). 如果$E\in\mathcal{S}$，那么E就是它本身的可测覆盖. 另一方面，如果$C\in\mathcal{S}$且C在$\mathcal{S}$中仅有的子集是它本身和$\varnothing$(但C可能是包含多于一个点的集合)，那么C是任何非空子集$E\subset C$的可测覆盖. 进一步地，如果存在集合$A\in\mathcal{S}$，满足$A\subset E\subset C$，$\mu(C\setminus A)=0$，则$\mu^*(E\setminus A)=\mu^*(C\setminus E)=0$. 由测度$\mu$为零的集合组成的集类定义为$\mathcal{N}(\mu):=\{F\subset X: \mu^*(F)=0\}$，则$\mathcal{N}(\mu)$中任意子集的可数并还在$\mathcal{N}(\mu)$中.

令$\mathcal{S}\vee\mathcal{N}(\mu)$为由$\mathcal{S}\cup\mathcal{N}(\mu)$生成的$\sigma$-代数，$\mathcal{S}_\mu^*:=\{E\subset X$: 对某个$B\in\mathcal{S}$，有$E\Delta B\in\mathcal{N}(\mu)\}$，由于$E\Delta B:=(E\setminus B)\cup(B\setminus E)$，从而称$E$与$B$几乎处处相等. $\mathcal{S}_\mu^*$是$\mathcal{S}$中几乎处处相等的集合组成的集类. 例如，如果$A\subset E\subset C$，$\mu(C\setminus A)=0$，那么$E\in\mathcal{S}_\mu^*$并且在定义中可取$B=A$或者C. 以上借助于零测

[101]

集扩张σ-代数的两种方法，都得到了相同的结论.

3.3.2 命题　*对任意测度空间$(X, \mathcal{S}, \mu)$及$\mathcal{S}$中几乎处处相等的集合组成的集类$\mathcal{S}_\mu^*$，有$\mathcal{S}_\mu^*=\mathcal{S}\vee\mathcal{N}(\mu)$，它是包括$\mathcal{S}$和$\mathcal{N}(\mu)$的最小$\sigma$-代数.*

证明　如果$E\Delta B\in\mathcal{N}(\mu)$且$B\in\mathcal{S}$，那么$E\setminus B\in\mathcal{N}(\mu)$，$B\setminus E\in\mathcal{N}(\mu)$，所以$E=(B\cup(E\setminus B))\setminus(B\setminus E)\in\mathcal{S}\vee\mathcal{N}(\mu)$，从而$\mathcal{S}_\mu^*\subset\mathcal{S}\vee\mathcal{N}(\mu)$.

反之，$\mathcal{S}\subset\mathcal{S}_\mu^*$，$\mathcal{N}(\mu)\subset\mathcal{S}_\mu^*$. 如果对所有的$n$，有$A_n\Delta B_n\in\mathcal{N}(\mu)$，那么

$$(\bigcup_n A_n)\Delta(\bigcup_n B_n)\subset\bigcup_n (A_n\Delta B_n)\in\mathcal{N}(\mu),$$

且$(X\setminus A_1)\Delta(X\setminus B_1)=A_1\Delta B_1\in\mathcal{N}(\mu)$.

所以$\mathcal{S}_\mu^*$是一个σ-代数，从而$\mathcal{S}\vee\mathcal{N}(\mu)\subset\mathcal{S}_\mu^*$. □

例如，令μ为整数集$\mathbb{Z}$上的测度，且对每个$k>0$，$\mu(k)>0$，对所有的$k\leqslant 0$，$\mu(k)=0$. 因此，$\mathcal{N}(\mu)$是非正整数集$-\mathbb{N}$的所有子集组成的环．令$\mathcal{S}$是所有$A\subset\mathbb{Z}$集合生成的σ-代数，对某个正整数集B，使得$A=B$，或者$A=-\mathbb{N}\cup B$. 则命题3.3.2中的两个σ-代数相等且都等于由$\mathbb{Z}$的所有子集生成的σ-代数$2^{\mathbb{Z}}$.

对给定的测度μ，下面的命题就说明了可以把μ扩张为测度$\bar{\mu}$，无论何时均可使得这两个集合关于μ几乎处处相等，并且μ至少被定义为它们中的一个，从而为使这两者都相等，定义$\bar{\mu}$如下：如果$A\Delta B\in\mathcal{N}(\mu)$，$B\in\mathcal{S}$，令$\bar{\mu}(A):=\mu(B)$，这里$A$可以是命题3.3.2中描述的$\sigma$-代数$\mathcal{S}\vee\mathcal{N}(\mu)$中的任意集合，其中$\bar{\mu}$是一个测度，具体描述如下.

3.3.3 命题　*对任意测度空间$(X,\mathcal{S},\mu)$，$\bar{\mu}$是σ-代数$\mathcal{S}\vee\mathcal{N}(\mu)$上明确定义的测度，且在$\mathcal{S}$上$\bar{\mu}$与$\mu$相等.*

证明　首先证$\bar{\mu}$是明确定义的测度．令$A\Delta B\in\mathcal{N}(\mu)$，$A\Delta C\in\mathcal{N}(\mu)$，$B$，$C\in\mathcal{S}$. 那么$B\Delta C=(A\Delta B)\Delta(A\Delta C)$.（为此，令$D$和$E$为任意集合，注意到，$1_{D\Delta E}=1_D+1_E$的加法是模2的，也就是说2可以由零代替．这一加法满足交换律和结合律．当其中的集合由示性函数所代替，Δ由$+$所代替时，代替后的方程就是我们希望得到的结果．）而从$B\Delta C\in\mathcal{N}(\mu)$，因此，$\mu(B)=\mu(C)$且$\bar{\mu}$是明确定义的测度.

假设$A_n\in\mathcal{S}\vee\mathcal{N}(\mu)$并且$A_n$是不相交的．取$B_n\in\mathcal{S}$，使得$C_n:=A_n\Delta B_n\in\mathcal{N}(\mu)$，则对$i\neq j$，$B_i\cap B_j\subset C_i\cup C_j$，从而$\mu(B_i\cap B_j)=0$. 令$D_n:=B_n\setminus\bigcup_{i<n}B_i$，$B:=\bigcup_{i\geqslant 1}B_i$，则对所有的$n$，$\mu(D_n)=\mu(B_n)$，$\mu(B)=\sum_{n\geqslant 1}\mu(D_n)=\sum_n\mu(B_n)$. 正如在命题3.3.2的证明中所表明的，如果$A:=\bigcup_{n\geqslant 1}A_n$，则$A\Delta B\in\mathcal{N}(\mu)$，因此，$\bar{\mu}(A)=\mu(B)=\sum_n\bar{\mu}(A_n)$并且$\bar{\mu}$是可数可加的．显然，在$\mathcal{S}$上$\bar{\mu}$与$\mu$相等. □

102

测度$\bar{\mu}$称为测度μ的*完备化*(completion).

在3.1节中我们是从代数学上的非负可数可加函数μ出发进行讨论的．令$\tilde{\mu}$是μ^*在$\mathcal{M}(\mu^*)$上的限制，那么它是一个测度且由引理3.1.6和引理3.1.8知，它是对测度μ的扩张．扩张的测度通常也记为μ. 为了使这个概念有意义，我们需要验证外测度μ^*不因这个扩张而有所改变.

3.3.4 命题　*在集合X中，对代数$\mathcal{A}$上的任意非负可数可加函数μ，在2^X上有$\tilde{\mu}^*=\mu^*$.*

证明　因为在$\mathcal{A}$上$\tilde{\mu}=\mu$（根据引理3.1.6），显然，$\tilde{\mu}^*\leqslant\mu^*$. 反之，对任意$A\subset X$，由定理3.3.1，取$C\in\mathcal{M}(\mu^*)$，使得$A\subset C$，$\tilde{\mu}^*(A)=\tilde{\mu}(C)$，则

$$\mu^*(A)\leqslant\mu^*(C)=\tilde{\mu}(C)=\tilde{\mu}^*(A).$$ □

测度空间$(X,\mathcal{S},\mu)$称为*有限的*当且仅当$\mu(X)<+\infty$，称为*σ-有限的*当且仅当存在集合序列$E_n\in\mathcal{S}$，使得$\mu(E_n)<\infty$且$X=\bigcup_n E_n$. 在这样的空间中，命题3.3.2中所描述的σ-代数与3.1节中所定义的σ-代数$\mathcal{M}(\mu^*)$相等（$\mathcal{M}(\mu^*)$中的元素就是所有可分解为对μ^*可加的集合）.

3.3.5 定理　*对任意σ-有穷测度空间$(X,\mathcal{S},\mu)$，$\mathcal{S}\vee\mathcal{N}(\mu)=\mathcal{M}(\mu^*)$.*

证明 令 $X=\bigcup_n E_n$，$\mu(E_n)<\infty$，$E\in\mathcal{M}(\mu^*)$，为证明 $E\in\mathcal{S}\vee\mathcal{N}(\mu)$，只需证明对所有的 n，有 $E\cap E_n\in\mathcal{S}\vee\mathcal{N}(\mu)$ 成立，因此，假设 μ 是有限的．由定理 3.3.1，设 A 是 E 的一个可测覆盖，那么 $\mu^*(E)=\mu(A)=\mu^*(A)=\mu^*(E)+\mu^*(A\setminus E)$，从而 $\mu^*(A\setminus E)=0$ 且 $E=A\setminus(A\setminus E)\in\mathcal{S}\vee\mathcal{N}(\mu)$，因此，$\mathcal{M}(\mu^*)\subset\mathcal{S}\vee\mathcal{N}(\mu)$．相反的包含关系可由引理 3.1.7、引理 3.1.8 和命题 3.1.9 得出． □

例：令 $\mathcal{S}:=\{\varnothing, X\}$，$\mu(\varnothing)=0$，$\mu(X)=+\infty$，只要 $\varnothing\neq E\subset X$，就有 $\mu^*(E)=+\infty$．因此，当 $\mathcal{S}\vee\mathcal{N}(\mu)=\{\varnothing, X\}$ 时，X 的所有子集都属于 $\mathcal{M}(\mu^*)$，如果 X 是包含多于一个点的集合，则 X 就是一个更小的 σ-代数．由此可知为什么在定理 3.3.5 中必须要求 σ-有限．

拓扑可以在任意一个子集(2.1 节中所述)上诱导出一个“相对”拓扑，类似地，测度可以在任意一个子集上诱导出一个“相对”测度，但这一子集可能是不可测的.

103

***3.3.6 定理** 设 $(X,\mathcal{S},\mu)$ 为一个测度空间，Y 为集合 X 的任意子集，X 具有有限外测度 μ^*，满足 $\mu^*(Y)<\infty$．$\mathcal{S}_Y$ 为所有集合 $A\cap Y$ 组成的集类，其中 $A\in\mathcal{S}$，且对每个 $B\in\mathcal{S}_Y$，令 $\nu(B):=\mu^*(B)$，则 $(Y,\mathcal{S}_Y,v)$ 是测度空间．

证明 首先，易证 $\mathcal{S}_Y$ 为 Y 的子集的 σ-代数．令 C 是 Y 的可测覆盖(根据定理 3.3.1)，设 $A_n\in\mathcal{S}$，满足 $A_n\cap Y(n=1,2,\cdots)$ 为不相交集合序列．为证 $v\left(\bigcup_n A_n\cap Y\right)=\sum_n v(A_n\cap Y)$，我们用 $A_n\cap C$ 代替 A_n，此时 $A_n\cap Y$ 不变．对每个 n，令 $D_n:=A_n\setminus\bigcup_{j<n}A_j$，那么 $D_n\in\mathcal{S}$. 因为 $A_n\cap Y$ 是不相交的，所以对所有的 n，有 $D_n\cap Y=A_n\cap Y$. 因此，我们可以用 D_n 代替 A_n，并假设 A_n 是不相交的且 $A_n\subset C$. 为证 $v(A_n\cap Y)=\mu(A_n)$，显然有 $v(A_n\cap Y)\leqslant\mu(A_n)$．令 F_n 是 $A_n\cap Y$ 的可测覆盖，则 $v(A_n\cap Y)=\mu(F_n)$．假设 $F_n\subset A_n$，如果 $\mu(F_n)<\mu(A_n)$，则 $A_n\setminus F_n\in\mathcal{S}$，它在 Y 中是不相交的且具有正测度，从而 C 可由集合 $C\setminus(A_n\setminus F_n)$ 代替，这与 C 是 Y 的可测覆盖矛盾．因此，对所有的 n，有 $v(A_n\cap Y)=\mu(A_n)$．令 $A:=\bigcup_n A_n$，那么 $\sum_n\mu(A_n)=\mu(A)$．借助于对每个 A_n 的相同讨论，我们有 $\mu(A\cap Y)=\mu(A)$，从而 v 是可数可加的． □

习题

在下面的习题中，$(X,\mathcal{S},\mu)$ 是可测空间．

1. 证明：由所有满足 $\mu(A)<\infty$ 的集合 $A\in\mathcal{S}$ 组成的集类是一个环．
2. (a) 证明：集合的对称差分满足结合律：对任意集合 A，B 和 C，有
$$A\Delta(B\Delta C)=(A\Delta B)\Delta C.$$
   (b) 如果 $A_1, A_2, \cdots, A_n$ 为任意集，$A:=A_1\Delta(A_2\Delta(A_3\Delta(\cdots(A_{n-1}\Delta A_n)\cdots)$，证明：$A$ 由 A_j 中的点组成，其中 j 为奇数．
3. 如果 $\bar{\mu}$ 是 μ 的完备化，证明：E 在 σ-代数中且在此 σ-代数上 $\bar{\mu}$ 是有定义的，当且仅当在 μ 的定义域内存在集合 A 和 C，使得 $A\subset E\subset C$ 且 $\mu(C\setminus A)=0$.
4. 证明或者反证：如果 $A\subset E\subset C$，$\mu(A)=\mu(C)$，那么 E 在完备化测度 $\bar{\mu}$ 的定义域内．[提示：如果 $\mu(A)=\infty$，会怎样?]
5. 设 μ 为[0, 1]上的测度，若 μ 在所有的第一范畴集(无处稠密集的可数并集)上为 0，在这类集合的所有补集上为 1，则 μ 的完备化测度是什么？

104

6. 对由 $\mu_*(A):=\sup\{\mu(B): B\subset A\}$ 定义的内测度 μ_*，证明：对任意集合 $A\subset X$，存在可测集 $B\subset A$，满足

$\mu(B)=\mu_*(A)$.

7. 如果$(X, \mathcal{S}, \mu)$是一个测度空间，A_1，A_2，…为 X 的任意子集，C_j 为 A_j 的可测覆盖，其中 $j=1, 2, \cdots$.
   (a)证明：$\bigcup_j C_j$ 是 $\bigcup_j A_j$ 的可测覆盖．
   (b)举例说明 $C_1\cap C_2$ 不一定是 $A_1\cap A_2$ 的可测覆盖．
8. 集类$\mathcal{R}$称为 σ-环(σ-ring)，当且仅当$\mathcal{R}$是环且$\mathcal{R}$中任意集合的可数并还在$\mathcal{R}$中．证明：如果$\mathcal{R}$是集合 X 的子集的 σ-环，则$\mathcal{R}\cup\{X\setminus A: A\in\mathcal{R}\}$是一个 σ-代数．如果 $X\notin\mathcal{R}$，则$\mathcal{R}$不是 σ-代数，但 μ 是一个由$\mathcal{R}$映射到$[0, \infty]$的可数可加函数，证明：对每个 $A\in\mathcal{R}$，令 $\mu(X\setminus A):=+\infty$，可使 μ 成为一个测度．
9. 由集合组成的集类$\mathcal{N}$称为遗传的(hereditary)，如果 $A\subset B\in\mathcal{N}$就有 $A\in\mathcal{N}$. 另外，如果对任意集合序列 $A_n\in\mathcal{N}$，有 $\bigcup_n A_n\in\mathcal{N}$，那么称$\mathcal{N}$为一个遗传 σ-环(hereditary σ-ring)．设$\mathcal{N}$为集合 X 的子集的遗传 σ-环．
   (a)证明：$\mathcal{N}$是 σ-环.
   (b)证明：$\mathcal{N}\vee\mathcal{S}$是由满足对某个 $B\in\mathcal{S}$，有 $A\Delta B\in\mathcal{N}$的所有集合 $A\subset X$ 组成的集类．
   (c)如果对所有的 $A\in\mathcal{N}\cap\mathcal{S}$，有 $\mu(A)=0$，证明：μ 可扩张为$\mathcal{S}\vee\mathcal{N}$上的测度且对所有的 $C\in\mathcal{N}$，有 $\mu(C)=0$.

3.4 勒贝格测度和不可测集

在前面部分，我们没有明确阐述$\mathbb{R}$ 的所有子集关于(定理 3.2.6 中定义的)勒贝格测度 λ 是否都是可测的．关于 λ 存在不可测集的证明需要选择公理．下面给出关于其他测度为不可测集的简单例子．首先，设 $X=\{0, 1\}$，$\mathcal{S}$为平凡的 σ-代数$\{\varnothing, X\}$，$\mu(X)=1$，那么 $\mu^*(\{0\})=\mu^*(\{1\})=1$，所以$\{0\}$和$\{1\}$关于$\mu^*$是不可测的．另一个较为常见的例子是，设 X 为任意不可数集且$\mathcal{S}:=\{A\subset X: A$ 或者 $X\setminus A$ 是可数的$\}$，如果 A 可数，则 $m(A)=0$，如果 A 的补集可数，则 $m(A)=1$，那么$\mathcal{S}$是 σ-代数且 m 是其上的测度．X 中任何具有不可数补集的不可数集 C 是不可测的，这是因为

$$m^*(C)+m^*(X\setminus C)=1+1=2\neq m^*(X)=m(X)=1.$$

[105]

称集合 $E\subset\mathbb{R}$ 为勒贝格可测的(Lebesgue measurable)，当且仅当 $E\in\mathcal{M}(\lambda^*)$．由引理 3.1.6 和引理 3.1.8 知，限制在勒贝格可测集的 σ-代数$\mathcal{M}(\lambda^*)$上的 λ^*是测度 λ 的扩张测度．这个测度仍称为 λ 测度或者勒贝格测度．$\mathcal{M}(\lambda^*)$通常是 λ 所定义的最大定义域．(这一定义域可扩张为较大的 σ-代数，但仅仅是以任意方式扩张的，见习题 3～5)．由命题 3.3.2 和定理 3.3.5 知，对每个勒贝格可测集 E，存在博雷尔集 B，满足 $\lambda(E\Delta B)=0$.

我们将继续证明，并不是所有的集合都是勒贝格可测的．一种可行的证明方法是(利用集合的势)证明一般集合比勒贝格可测集合多．已知 c 是集$[0, 1]$或者$\mathbb{R}$ 的势，换句话说，称一个集合 X 的势为 c，也就是说存在一个从$[0, 1]$映射到 X 的一一对应函数．在同种意义下，$2^{\mathbb{R}}$ 的势记为 2^c. 这一方法仅是建议的，但不一定是可操作的，因为$\mathbb{R}$ 中的勒贝格可测集与$\mathbb{R}$ 中的集合一样多．

***3.4.1 命题**　2^c 是勒贝格可测集．

证明　令 C 为康托尔集，

$$C:=\Big\{\sum_{n\geq 1}x_n/3^n: 对所有的\ n, x_n=0\ 或\ 2\Big\}.$$

对每个 $N=1, 2, 3, \cdots$，令 $C_N:=\Big\{\sum_{n\geq 1}x_n/3^n$：对所有的 n，有 $x_n=0$，1 或 2，且对 $n\leq N$，有 $x_n\neq 1\Big\}$.
从而 C_1 是单位区间$[0, 1]$中去掉开区间$(1/3, 2/3)$的区间．那么为得到 C_2，只需在剩余的两个区间中分别去掉中间的三分之一区间$(1/9, 2/9)$和$(7/9, 8/9)$．将这一过程反复进行 N 次可得到 C_N. 因

而 $C_1 \supset C_2 \supset \cdots \supset C_N \supset \cdots$，$\bigcap_{N\geq 1} C_N = C$，对所有的 N，有 $\lambda(C_N) = (2/3)^N$，所以 $\lambda(C) = 0$. 另一方面，由于 C 具有势 c，所以 C 有 2^c 个子集且都是测度为 0 的勒贝格测度．因此，$\mathbb{R}$ 有 2^c 个子集，确切地说，它有 2^c 个勒贝格可测集(根据等价定理 1.4.1). □

大家可能会问是否存在正测度的勒贝格可测集 A，其完备化还具有正测度且 A 可以在线上均匀伸展，在这一意义下，对满足 $0<r<1$ 的某个常数 r 及对所有满足 $0<\lambda(J)<+\infty$ 的区间 J，有比率 $\lambda(A\cap J)/\lambda(J) = r$. 由此证实了这样的集合 A 是不存在的，对于任意正测度集，其完备化仍具有正测度，一些比率一定接近于 0，其他的比率一定接近于 1. 特别地，有如下命题.

106

3.4.2 命题 设 E 是 $\mathbb{R}$ 中满足 $\lambda(E)>0$ 的任意勒贝格测度，那么对任意 $\varepsilon>0$，存在一个有限非平凡区间 $J(J=[a, b],\ -\infty<a<b<+\infty)$，使得 $\lambda(E\cap J)>(1-\varepsilon)\lambda(J)$.

证明 假设 $\lambda(E)<\infty$，$\varepsilon<1$，取有限区间 $J_n := (a_n, b_n]$，$a_n<b_n$，使得 $E\subset \bigcup_n J_n$，$\sum_{n\geq 1}\lambda(J_n) \leq \lambda(E)/(1-\varepsilon)$，那么 $\lambda(E)\leq \sum_{n\geq 1}\lambda(J_n\cap E)$ 且对某个 n，有 $(1-\varepsilon)\lambda(J_n)\leq\lambda(J_n\cap E)$. □

一个集合 E 可具有正的勒贝格测度，事实上，它的补集具有 0 测度，包含任意区间长度为正的 E 是不存在的(例如，取无理数集). 但是如果我们取所有正测度集的元素之间的差分，下面的命题说明可以得到一个包含 0 周围的非平凡区间的集合．关于可测集构造的这一命题将用来证明不可测集是存在的．

3.4.3 命题 如果 E 为任意勒贝格可测集，且 $\lambda(E)>0$，那么对某个 $\varepsilon>0$，

$$E - E := \{x-y: x, y\in E\} \supset [-\varepsilon, \varepsilon].$$

证明 由命题 3.4.2，取区间 J，使得 $\lambda(E\cap J)>3\lambda(J)/4$. 令 $\varepsilon := \lambda(J)/2$. 对任意的 $C\subset\mathbb{R}$ 和 $x\in\mathbb{R}$，令 $C+x := \{y+x:\ y\in C\}$. 则如果 $|x|\leq\varepsilon$，有

$$(E\cap J)\cup((E\cap J)+x)\subset J\cup(J+x),$$
$$\lambda(J\cup(J+x))\leq 3\lambda(J)/2,$$

而

$$\lambda((E\cap J)+x) = \lambda(E\cap J),$$

所以

$$((E\cap J)+x)\cap(E\cap J)\neq\varnothing,$$
$$x\in(E\cap J)-(E\cap J)\subset E-E.$$ □

下面将给出本节关于不可测集存在性的一个主要定理. 特别地，在 $[0, 1]$ 上外测度为 1 的集合 E，所以 E 在整个区间上是稠密的，它的补集同样也是稠密的.

3.4.4 定理 假定(通常的)选择公理成立，存在一个集合 $E\subset\mathbb{R}$，E 不是勒贝格可测的．事实上，存在一个集合 $E\subset I := [0, 1]$，满足 $\lambda^*(E) = \lambda^*(I\setminus E) = 1$.

107

证明 记 $\mathbb{Z}$ 为所有整数(正整数、负整数或者 0)组成的集合．令 α 是一固定的无理数，设 $\alpha = 2^{1/2}$. 令 G 是 $\mathbb{R}$ 的可加子群：$G := \mathbb{Z} + \mathbb{Z}\alpha := \{m+n\alpha:\ m, n\in\mathbb{Z}\}$. 令 H 为子群，$H := \{2m+n\alpha:\ m, n\in\mathbb{Z}\}$. 首先证明 G 在 $\mathbb{R}$ 中稠密．令 $c := \inf\{g:\ g\in G,\ g>0\}$. 如果 $c=0$，令 $0<g_n<1/n$，$g_n\in G$. 那么 $G\supset\{mg_n:\ m\in\mathbb{Z},\ n=1, 2, \cdots\}$ 且 G 是稠密集．如果 $c>0$ 且 $g_n\downarrow c$，$g_n\in G$，$g_n>c$，那么 $g_n - g_{n+1}>0$，且属于 G 并收敛到 0，所以 $c=0$，矛盾．因此，$c\in G$ 且 $G=\{mc:\ m\in\mathbb{Z}\}$，而 α 为无理数，矛盾．同样，H 和 $H+1$ 是稠密的．由于陪集 $G+y(y\in\mathbb{R})$ 是不相交或者恒等的，根据选

择公理，令 C 是恰好包含每一陪集的一个元素的集合．令 $X := C + H$，那么$\mathbb{R} \setminus X = C + H + 1$. 又 $(X - X) \cap (H + 1) = \varnothing$，因为 $H + 1$ 是稠密的，由命题3.4.3知，X 不包括任意具有正勒贝格测度的可测集．令 $E := X \cap I$，那么 $\lambda^*(I \setminus E) = 1$. 同样，$(\mathbb{R} \setminus X) - (\mathbb{R} \setminus X) = (C + H + 1) - (C + H + 1) = (C + H) - (C + H)$与 $H + 1$ 是不相交的，所以$\lambda^*(E) = 1$. □

勒贝格测度没有被定义在 I 的所有子集上，但是它能否作为一个可数可加测度扩张到 I 的所有子集上呢？至少，如果这一连续统假设满足的话，答案是否定的(附录 C).

习题

1. 设 C 是一个勒贝格可测集，对$\mathbb{R}$ 的一个稠密子集中的所有 x，有 $\lambda(E\Delta(E+X)) = 0$，证明：$\lambda(E) = 0$ 或 $\lambda(\mathbb{R} \setminus E) = 0$.
2. 证明：在$[0, 1]$中，对所有的 k，存在满足 $\lambda^*(A_k) = 1$，$\bigcap_k A_k = \varnothing$的集合 $A_1 \supset A_2 \supset \cdots$.

   [提示：C 和 α 同定理 3.4.4 的证明中一致，设 $B_k := \{m + n\alpha: m, n \in \mathbb{Z}, |m| \geqslant 2k\}$，$A_k := (C + \{m + n\alpha: m, n \in \mathbb{Z}, |m| \geqslant k\}) \cap [0, 1]$，证明 B_k 在$\mathbb{R}$ 中稠密且$([0, 1] \setminus A_k) - ([0, 1] \setminus A_k)$与 B_k 是不相交的.]
3. 如果$\mathcal{S}$是 X 的子集的σ-代数，E 是 X 的任意子集，证明：由$\mathcal{S} \cup \{E\}$生成的σ-代数是所有形如$(A \cap E) \cup (B \setminus E)$的集合组成的集类，其中 A，$B \in \mathcal{S}$.
4. 证明：对任意有限测度空间$(X, \mathcal{S}, \mu)$和任意集合 $E \subset X$，测度 μ 总是可以扩张为 σ-代数$\mathcal{T} := \mathcal{S} \vee \{E\}$上的测度 ρ. [提示：在习题 3 所给出的形式中，设$\rho(A \cap E) = \mu^*(A \cap E)$，$\rho(B \setminus E) := \mu_*(B \setminus E) := \sup\{\mu(C): C \in \mathcal{S}, C \subset B \setminus E\}$. 提示：参见定理 3.3.6.]
5. 参考习题 4，证明：可把测度 μ 扩张为定义在 E 上的测度 α，其中 $\alpha(E)$可能取区间$[\mu_*(E), \mu^*(E)]$上的任意值.

108

6. 如果$(X, \mathcal{S}, \mu)$是一个测度空间，A_n 为$\mathcal{S}$中满足 $\mu(A_1) < \infty$，$A_n \downarrow A$ 的集合，即 $A_1 \supset A_2 \supset \cdots$，$\bigcap_n A_n = A$，证明：$\lim_{n\to\infty} \mu(A_n) = \mu(A)$. [提示：集合 A_n / A_{n+1}是不相交的.]
7. 设 μ 为一个定义在由$[0, 1]$的子集所组成的博雷尔 σ-代数上的有限测度，对每一单点 p，满足 $\mu(\{p\}) = 0$. 设 $\varepsilon > 0$.

   (a)证明：对任意 p，存在一个包含 p 的开区间 J，满足 $\mu(J) < \varepsilon$.

   (b)证明：存在一个满足 $\mu(U) < \varepsilon$ 的稠密开集合 U.
8. 对每个如同定理 2.5.2 之后所定义的第一范畴博雷尔集合 A，A 属于 $I := [0, 1]$. 设 $\mu(A) = 0$，$\mu(I \setminus A) = 1$. 证明：μ 不能被扩张为博雷尔 σ-代数上的测度. [提示：用习题 7.]
9. 如果$(X, \mathcal{S}, \mu)$是一个测度空间，$\{E_n\}$是 X 的一个子集序列，证明：μ 总能被扩张为博雷尔 σ-代数上的一个测度，且此博雷尔 σ-代数包含 $E_1, \cdots, E_n$ 但并不一定包含所有 E_n. [提示：利用习题 4 和习题 8，其中$\{E_n\}$是$[0, 1]$拓扑的一组基.]
10. 如果 A_k 同习题 2，满足 $A_k \downarrow \varnothing$和 $\lambda^*(A_k) \equiv 1$. 对每个博雷尔集 $B \subset [0, 1]$，设 $P_n(B \cap A_n) := \lambda^*(B \cap A_n) = \lambda(B)$. 证明：$P_n$ 是 A_n 的子集的一个 σ-代数上的可数可加测度，且对无限多的 n，A_{n+1}对在 A_n 中的 P_n 是不可测的. [提示：利用定理 3.3.6.]

*3.5　原子测度和非原子测度

如果$(X, \mathcal{S}, \mu)$是一个测度空间，集合 $A \in \mathcal{S}$称为μ 的一个原子(atom)，当且仅当$0 < \mu(A) < \infty$ 且对每个 $C \subset A$，$C \in \mathcal{S}$，或者 $\mu(C) = 0$ 或者 $\mu(C) = \mu(A)$. 没有任何原子的测度称为非原子的(nonatomic).

原子的主要实例是具有正有限测度的单元素集$\{x\}$. 一个正的有限测度集合是一个原子，如果它的可测子集仅仅是它本身和$\varnothing$. 这里指非平凡原子：设 X 是不可数集，$\mathcal{S}$是集合 A 的集类，A 或者可数，满足$\mu(A)=0$，或者具有可数补集，满足$\mu(A)=1$. 那么μ 是一个测度，X 是一个原子. 另一方面，勒贝格测度是非原子的(证明留作习题).

一个测度空间$(X, \mathcal{S}, \mu)$，或测度μ 称为*纯原子的*(purely atomic)，当且仅当存在由μ 的原子组成的集类$\mathcal{C}$，使得对每个 $A\in\mathcal{S}$，$\mu(A)$是对所有的 $C\in\mathcal{C}$，满足$\mu(A\cap C)=\mu(C)$的数$\mu(C)$的和.(对任意的非负实数a_C，$\sum\{a_C: C\in\mathcal{C}\}$定义为$\mathcal{C}$上所有有限子集的和的上确界.)纯原子测度的主要例子是，存在一个函数$f\geqslant 0$，使得$\mu(A)=\sum\{f(x): x\in A\}$. 满足$f\equiv 1$ 的计数测度是纯原子的. 在$\mathbb{R}$上，研究最多的纯原子测度集中在对于某个$c_n\geqslant 0$，满足$\mu(A)=\sum_n c_n 1_A(x_n)$的可数集$\{x_n\}_{n\geqslant 1}$.

109

人们对具有无限测度的集合并不感兴趣的，它可能会带来一些技术困难，除非它们有任意大有限测度的子集，对σ-有限测度和下面较为一般的个别测度，这也成立. 测度空间$(X, \mathcal{S}, \mu)$称为*可局部化的*(localizable)，当且仅当存在由具有有限测度的不相交可测集组成的集类$\mathcal{A}$，这些子集的并为 X，使得对每个 $B\subset X$，B 是可测的当且仅当对所有 $C\in\mathcal{A}$，有 $B\cap C\in\mathcal{S}$，从而有$\mu(B)=\sum_{C\in\mathcal{A}}\mu(B\cap C)$. 最有用的可局部化测度是满足$\mathcal{A}$可数的$\sigma$-有限测度. 在可能不可数集合上的计数测度提供了其他的例子.

在实际中考虑较多的测度或者是纯原子的或者是非原子的，但总可以把一个纯原子的有限测度添加到一个非原子测度中，从而得到一个测度，对这个测度，下面的分解是非平凡的.

3.5.1 定理 *设$(X, \mathcal{S}, \mu)$是一个可局部化空间，那么存在$\mathcal{S}$上的测度ν和ρ，使得$\mu=\nu+\rho$，其中ν是纯原子的，ρ是非纯原子的.*

定理 3.5.1 的证明略，证明的细节留作习题 1~7. 首先，可以归约为有限测度空间的情形. 设$\mathcal{C}$为μ 的所有原子组成的集类，对原子 A，B，定义关系 $A\approx B$ 当且仅当$\mu(A\cap B)=\mu(A)$. 这是个等价关系. 设 I 是所有等价类的集合且对每个 $i\in I$，在等价类 i 中选择一个原子 C_i. 令 $\nu(A)=\sum_{i\in I}\mu(A\cap C_i)$，且$\rho=\mu-\nu$.

习题

1. 证明：如果定理 3.5.1 对有限测度空间成立，那么它对所有可局部化测度空间也成立.
2. 证明：在可局部化测度μ 的定义中，或者$\mu\equiv 0$，或者选择$\mathcal{A}$，使得对所有的 $C\in\mathcal{A}$，$\mu(C)>0$.
3. 设μ 为有限的，证明：$\approx$是一个等价关系.
4. 设μ 为有限的，在这种情况下，如果 A，B 是两个不等价的原子，证明：$\mu(A\cap B)=0$.
5. 证明：如上所定义的ν是一个纯原子测度且$\nu\leqslant\mu$.
6. 证明：对任意测度 $\nu\leqslant\mu$，存在一个满足 $\mu=\nu+\rho$ 的测度 ρ. [提示：对μ 是有限的容易证明，但若$\rho=\mu-\nu$，ρ 对满足$\mu(A)=\nu(A)=\infty$ 的集合 A 没有定义，在这样的集合上，设$\rho(A):=\sup\{(\mu-\nu)(B): \nu(B)<\infty, B\subset A\}$.]

110

7. 设μ 和ν 同习题 5~6 中的，证明：ρ 是非原子的.
8. 对给定的测度空间$(X, \mathcal{S}, \mu)$，可测集 A 称为具有*纯无限测度*(purely infinite measure)，当且仅当$\mu(A)=+\infty$且对每个可测集 $B\subset A$，或者$\mu(B)=0$，或者$\mu(B)=+\infty$. 称这样的两个集合 A 和 C 是等价的，当且仅当$\mu(A\Delta C)=0$. 给出一个测度空间与两个纯无限集合 A 和 C 的例子，其中集合 A 和 C 是不等价的，但$\mu(A\cap C)=+\infty$.
9. 证明：勒贝格测度是非原子的.

10. 设 X 是一个可数集，证明：X 上的任意测度是纯原子的．
11. 如果$(X,\ \mathcal{S},\ \mu)$是一个测度空间，$\mu(X)=1$ 且 μ 是非原子的，证明：μ 的取值范围是整个区间$[0,\ 1]$．[提示：首先证明对某个 C，$1/3\leq\mu(C)\leq 2/3$．否则，证明在集合 B 上取到的最大值 $s<1/3$，且 B 的补集包含一个原子．将 μ 限制在 C 和 C 的补集上重复讨论可得到中间测度的集合，反复进行即可得到 μ 值的一个稠密集．]

注释

3.1 节　Jordan(1892)定义一个集合是"可测的"，如果它的拓扑边界测度为 0. 因此，在若尔当所定义的意义下，有理数集$\mathbb{Q}$是不可测的．Borel(1895，1898)证明了区间的长度可扩张为由区间生成的 σ-代数上的可数可加函数，它包含所有可数集．后来，σ-代数由他命名．Fréchet(1965)写了一本关于博雷尔的传记论文集．博雷尔写了 35 本书和 250 多篇论文，他的数学论文收集在 Borel(1972)中．从 1924 年到 1936 年，他还被选举为法国国会议员，1925 年他成为内阁成员，主管海军．博雷尔较少的一些与数学哲学和自然哲学联系较多的研究论文收集在 Borel(1967)中．

Hawkins(1970，第 4 章)回顾了关于可测集的发展史．Radon(1913)可能是第一个在一般空间(维数超过$\mathbb{R}^k$)上定义测度的人．Caratheodory(1918)显然是第一个定义外测度 μ^* 和可测集的集类$\mathcal{M}(\mu^*)$的人，并证明了$\mathcal{M}(\mu^*)$是一个 σ-代数且 μ^* 限制在$\mathcal{M}(\mu^*)$上是测度．

为什么要假设可数可加呢？对任意不可数闭区间的并，其长度是不可加的，例如，$[0,\ 1]$是长度为 0 的 c 个单元素集$\{x\}=[x,\ x]$的并．因而在这种不可数区间的并上，其长度可加性似乎是太强的假设．另一方面，有限可加性在处理如 3.1 节最后部分的一些反常问题时显得较弱．当前，通常所定义的概率是遵循 Kolmogorov(1933)的定义，即作为集合 X 子集的 σ-代数$\mathcal{S}$上的一个(可数可加)测度且满足 $P(X)=1$. 在关于概率的不多的研究中，研究有限可加概率"测度"的一部有名的著作是 Dubins 和 Savage(1965)．

111

3.2 节　半环的概念在 von Neumann(1940—1941，p. 79)的一些演讲记录中以不同的"$\mathcal{D}$型的"集类名字被提及．

(网格)黎曼积分的定义，对有限区间$[a,\ b]$上的连续函数 f，$a=x_0\leq y_1\leq x_1\leq y_2\leq\cdots\leq x_n=b$，当$\max_i(x_i-x_{i-1})\to 0$，$n\to\infty$ 时，和的极限

$$\sum_{i=1}^{n} f(y_i)(x_i-x_{i-1})$$

存在．Stieltjes(1894，p. 68—76)对函数 G 定义了与此类似的积分 $\int f\mathrm{d}G$，在上述和中用 $G(x_i)-G(x_{i-1})$ 代替 x_i-x_{i-1} 即可．积分结果称为黎曼–斯蒂阶(Riemann-Stieltjes)积分．测度 μ_G 称为勒贝格–斯蒂阶测度，尽管这一测度在 1894 年没有被定义．

3.4 节　根据勒贝格(1907，p. 212)，Vitali(1905)可能第一个证明了非勒贝格可测集的存在性．Van Vleck(1908)证明了定理 3.4.4 中的集合 E 是存在的．Solovay(1970)证明了原子的选择是不可缺少的，同时证明可数多个原子的选择是不充分的，因此，为了获得一个不可测集有必要进行许多不可数选择．(他的这一结论的精确论述需要的条件涉及太多技巧，因此在这里没有给出．)

3.5 节　Segal(1951)定义并研究了可局部化空间．

112

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到．

*Borel, Émile (1895). Sur quelques points de la théorie des fonctions. *Ann. Ecole Normale Sup.* (Ser. 3) *12*: 9–55, = *Œuvres*, I (CNRS, Paris, 1972), pp. 239–285.

——— (1898). *Leçons sur la théorie des fonctions.* Gauthier-Villars, Paris.

——— (1967). *Émile Borel: philosophe et homme d'action.* Selected, with a Préface, by Maurice Fréchet. Gauthier-Villars, Paris.

*——— (1972). *Œuvres*, 4 vols. Editions du Centre National de la Recherche Scientifique, Paris.

Caratheodory, Constantin (1918). *Vorlesungen über reelle Funktionen.* Teubner, Leipzig. 2d ed., 1927.

Dubins, Lester E., and Leonard Jimmie Savage (1965). *How to Gamble If You Must.* McGraw-Hill, New York.

Fréchet, Maurice (1965). *La vie et l'oeuvre d'Émile Borel.* L'Enseignement Math., Genève, also published in *L'Enseignement Math.* (Ser. 2) *11* (1965) 1–97.

Hawkins, Thomas (1970). *Lebesgue's Theory of Integration: Its Origins and Development.* Univ. Wisconsin Press.

Jordan, Camille (1892). Remarques sur les intégrales définies. *J. Math. pures appl.* (Ser. 4) *8*: 69–99.

Kolmogoroff, Andrei N. [Kolmogorov, A. N.] (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin. Published in English as *Foundations of the Theory of Probability*, 2d ed., ed. Nathan Morrison. Chelsea, New York, 1956.

Lebesgue, Henri (1907). Contribution a l'étude des correspondances de M. Zermelo. *Bull. Soc. Math. France 35*: 202–212.

von Neumann, John (1940–1941). Lectures on invariant measures. Unpublished lecture notes, Institute for Advanced Study, Princeton.

*Radon, Johann (1913). Theorie und Anwendungen der absolut additiven Mengenfunktionen. *Wien Akad. Sitzungsber. 122*: 1295–1438.

Segal, Irving Ezra (1951). Equivalences of measure spaces. *Amer. J. Math. 73*: 275–313.

Solovay, Robert M. (1970). A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. Math.* (Ser. 2) *92*: 1–56.

Stieltjes, Thomas Jan (1894). Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse* (Ser. 1) *8*: 1–122 and *9* (1895): 1–47, = *Oeuvres complètes*, II (Nordhoff, Groningen, 1918), pp. 402–566.

van Vleck, Edward B. (1908). On non-measurable sets of points with an example. *Trans. Amer. Math. Soc. 9*: 237–244.

*Vitali, Giuseppe (1905). *Sul problema della misura dei gruppi di punti di una retta.* Gamberini e Parmeggiani, Bologna.

第4章　积　　分

19 世纪的经典黎曼积分在处理某些函数时陷入了困境．例如：

(i)求函数 $x^{-1/2}$从 0 到 1 的积分，黎曼积分本身不适用．当 $\varepsilon\downarrow 0$ 时，必须对从 ε 到 1 的黎曼积分取极限.

(ii)同样，黎曼积分缺乏一些完备性．例如，如果$[0, 1]$上的函数f_n是连续的且对所有的 n 以及 x，有 $|f_n(x)|\leqslant 1$，当对所有的 x，$f_n(x)$收敛到某个$f(x)$时，黎曼积分 $\int_0^1 f_n(x)\,\mathrm{d}x$ 总是收敛的，但是黎曼积分 $\int_0^1 f(x)\,\mathrm{d}x$ 可能没有定义．

本章中定义和研究的勒贝格积分将使得 $\int_0^1 x^{-1/2}\,\mathrm{d}x$ 在任何时候都有定义，包括霍赫(hoc)极限过程，而且在(ii)中，$\int_0^1 f(x)\,\mathrm{d}x$ 将一直定义为勒贝格积分，而且当黎曼可积函数的勒贝格积分等于黎曼积分时，它也是积分序列 f_n 的极限．

勒贝格积分也适用于比$\mathbb{R}$ 更一般的空间上的函数，并基于更一般的测度．

4.1　简单函数

一个可测空间(measurable space)是一个序对$(X, \mathcal{S})$，其中 X 是一个集合，$\mathcal{S}$是一个由 X 的子集构成的 σ-代数．那么 X 上的一个简单函数(simple function)是任意有限和

4.1.1
$$f = \sum_i a_i 1_{B(i)},\quad \text{其中 } a_i\in\mathbb{R},\quad B(i)\in\mathcal{S}.$$

如果 μ 是$\mathcal{S}$上的一个测度，则称 f 为 μ-简单的(μ-simple)，当且仅当它是简单的而且能表示成式 4.1.1 的形式，其中对所有的 i，$\mu(B(i))<\infty$．(如果 $\mu(X)=+\infty$，那么对 $B(2)=B(2)=X$，$a_1=1$和 $a_2=-1$，有 $0=a_1 1_{B(1)}+a_2 1_{B(2)}$，但是 0 是一个 μ-简单函数．因此 μ-简单的定义只要求存在有限测度的 $B(i)$和 a_i，使得式 4.1.1 成立，并不要求这个 $B(i)$必须具有有限测度．)$\mathbb{R}$ 上的有关简单函数的例子是阶梯函数，其中每个 $B(i)$是一个有限区间．　[114]

集合 $B(1)$，…，$B(n)$的任意有限集类生成一个代数$\mathcal{A}$. 一个非空集合 A 称为代数$\mathcal{A}$的原子，当且仅当对所有的 $C\in\mathcal{A}$，或者 $A\subset C$ 或者 $A\cap C=\varnothing$时，$A\in\mathcal{A}$. 例如，如果 $X=\{1, 2, 3, 4\}$，$B(1)=\{1, 2\}$且 $B(2)=\{1, 3\}$，那么这两个集合生成一个由 X 的所有子集构成的代数，其中 X 的原子当然是单元素集$\{1\}$，$\{2\}$，$\{3\}$和$\{4\}$.

4.1.2 命题　设 X 为任意集，$B(1)$，…，$B(n)$为 X 的任意子集，$\mathcal{A}$是 X 的包含 $B(i)$$(i=1, \cdots, n)$的子集的最小 σ-代数，$\mathcal{C}$是所有交 $\bigcap_{1\leqslant i\leqslant n} C(i)$ 的集类，其中对每个 i，或者 $C(i)=B(i)$或者 $C(i)=X\setminus B(i)$，那么$\mathcal{C}\setminus\{\varnothing\}$是$\mathcal{A}$的所有原子的集合，而且$\mathcal{A}$中每个集都是它所包含的原子的并.

证明　$\mathcal{C}$中任意两个元素都是不相交的(对某个 i，一个包含在 $B(i)$中，另一个包含在 $X\setminus B(i)$中). $\mathcal{C}$ 的并是 X，因此，$\mathcal{C}$中元素的所有并的集合是一个代数$\mathcal{B}$. 每个 $B(i)$是$\mathcal{C}$中所有交的并．从而 $B(i)\in\mathcal{B}$且$\mathcal{A}\subset\mathcal{B}$. 显然$\mathcal{B}\subset\mathcal{A}$，所以$\mathcal{A}=\mathcal{B}$. 因此，$\mathcal{C}$中每个非空集是$\mathcal{A}$的一个原子．两个或更多个不

同原子的并不是一个原子，所以$\mathcal{C}\setminus\{\varnothing\}$是由$\mathcal{A}$的所有原子组成的集合，其余的结论也可得到． □

现在，任意的简单函数f能表示成$\sum_{1\leqslant j\leqslant M} b_j 1_{A(j)}$，其中$A(j)$是代数$\mathcal{A}$的不相交原子，$\mathcal{A}$是由式4.1.1中的$B(i)$生成的，根据命题4.1.2，有$M\leqslant 2^n$．因此，在式4.1.1中我们可以假定$B(i)$是不交的．那么，如果对所有的$x$，$f(x)\geqslant 0$，则对所有的$i$，有$a_i\geqslant 0$．例如：

$$3\cdot 1_{[1,3]}+2\cdot 1_{[2,4]}=3\cdot 1_{[1,2)}+5\cdot 1_{[2,3]}+2\cdot 1_{(3,4]}.$$

如果$(X,\mathcal{S},\mu)$是任意的测度空间，f是X上任意的简单函数，正如在式4.1.1中对所有的i，$a_i\geqslant 0$，f关于μ的积分定义为

4.1.3
$$\int f\mathrm{d}\mu := \sum_i a_i\mu(B(i))\in[0,\infty],$$

其中$0\cdot\infty$理解为0．首先，我们必须证明如下命题．

115

4.1.4 命题 对任意非负简单函数f，$\int f\mathrm{d}\mu$是明确定义的．

证明 假设$f=\sum_{i\in F} a_i 1_{E(i)}=\sum_{j\in G} b_j 1_{H(j)}$，$E_i:=E(i)$，$H_j:=H(j)$，其中$a_i$和$b_j$都是非负的，$F$和$G$是有限的，$E(i)$，$H(j)\in\mathcal{S}$，那么我们可以假设$H(j)$是由$H(j)$和$E(i)$生成的代数的不相交原子，在这种情况下，$b_j=\sum\{a_i:E_i\supset H_j\}$．因此，

$$\sum_j b_j\mu(H(j))=\sum_j \mu(H(j))\sum\{a_i:E_i\supset H_j\}$$
$$=\sum a_i\sum\{\mu(H_j):H_j\subset E_i\}=\sum a_i\mu(E_i).$$
□

如果f和g是简单的，那么显然$f+g$，fg，$\max(f,g)$和$\min(f,g)$都是简单的．从式4.1.3和命题4.1.4知，如果f和g是非负简单函数，c是一个常数且$c>0$，则$\int(f+g)\mathrm{d}\mu=\int f\mathrm{d}\mu+\int g\mathrm{d}\mu$且$\int cf\mathrm{d}\mu=c\int f\mathrm{d}\mu$．如果$0\leqslant f\leqslant g$，也就是对所有的$x$，$0\leqslant f(x)\leqslant g(x)$，那么$\int f\mathrm{d}\mu\leqslant\int g\mathrm{d}\mu$．对$E\in\mathcal{S}$，令$\int_E f\mathrm{d}\mu:=\int f1_E\mathrm{d}\mu$．例如，此时$\int_E 1_A\mathrm{d}\mu=\mu(A\cap E)$．

如果$(X,\mathcal{S})$和$(Y,\mathcal{B})$是可测空间，f是从X映射到Y的简单函数，那么f称为是*可测的*，当且仅当对所有的$B\in\mathcal{B}$，$f^{-1}(B)\in\mathcal{S}$．例如，如果$X=Y$，f是恒等函数，可测性意味着$\mathcal{B}\subset\mathcal{S}$．一般地，对于可测性，域空间上的$\sigma$-代数需要足够大，而且\或者值域空间上的$\sigma$-代数$\mathcal{B}$不太大．如果$Y=\mathbb{R}$或$[-\infty,\infty]$，那么对$Y$中的具有可测性的函数而言，$\sigma$-代数$\mathcal{B}$将(除非另有说明)由所有(有界或无界)区间或开集生成的博雷尔σ-代数．现在给定任意测度空间$(X,\mathcal{S},\mu)$和任意从X映射到$[0,+\infty]$的可测函数，定义

$$\int f\mathrm{d}\mu := \sup\left\{\int g\mathrm{d}\mu : 0\leqslant g\leqslant f, g\text{为简单函数}\right\}.$$

对$a_n\in[-\infty,\infty]$，$a_n\uparrow$意味着对所有的n，$a_n\leqslant a_{n+1}$．那么$a_n\uparrow a$也就是$a_n\to a$．如果$a=+\infty$，这意味着对所有的$M<\infty$，存在$K<\infty$，使得对所有的$n>K$，$a_n>M$．下面的命题给出了一个便利的方法，这种方法表明非负可测函数序列的极限上界比非负可测函数序列的极限积分的上界更具一般性．

4.1.5 命题 对任意可测的$f\geqslant 0$，存在简单函数f_n，它满足$0\leqslant f_n\uparrow f$，也就是对所有的x，

116

$0\leqslant f_n(x)\uparrow f(x)$．对任意这样的序列$f_n$，$\int f_n\mathrm{d}\mu\uparrow\int f\mathrm{d}\mu$．

证明　对 $n=1, 2, \cdots$，和 $j=1, 2, \cdots, 2^n n-1$，令 $E_{nj}:=f^{-1}((j/2^n, (j+1)/2^n])$，$E_n:=f^{-1}((n, \infty])$. 令

$$f_n := n1_{E_n} + \sum_{j=1}^{2^n n-1} j1_{E_{nj}}/2^n.$$

在图 4-1 中，当在$[0, \infty]$上 $f(x)=x$ 时，$g_n=f_n$，所以对 $0\leqslant x\leqslant n$，g_n 是具有宽度和高度为$1/2^n$的阶梯函数，且对 $x>n$，$g_n(x)\equiv n$. 那么对一般的 $f\geqslant 0$，有 $f_n=g_n\circ f$. 现在 $E_{nj}=E_{n+1,2j}\cup E_{n+1,2j+1}$，所以在 E_{nj}上，有 $f_n(x)=j/2^n=2j/2^{n+1}<(2j+1)/2^{n+1}$，其中 $f_{n+1}(x)$是后两项中的一项．因此，$f_n(x)\leqslant f_{n+1}(x)$. 在 E_n 上，$f_n(x)=n\leqslant f_{n+1}(x)$. 对于不在 E_n 或任意 E_{nj} 中的点 x，$f_n(x)=0\leqslant f_{n+1}(x)$. 因此在任意点处 $f_n\leqslant f_{n+1}$. 如果 $f(x)=+\infty$，那么对所有的 n，$f_n(x)=n$. 否则对某个 $m\in\mathbb{N}$，$f(x)<m$. 那么对 $n\geqslant m$，$f_n(x)\geqslant f(x)-1/2^n$，所以对所有的$x$，$f_n(x)\to f(x)$.

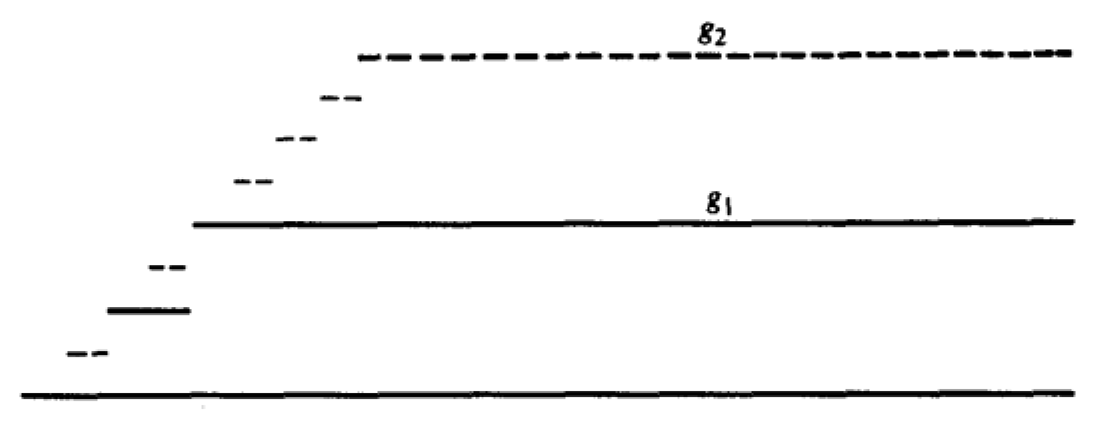


图　4-1

令 g 是任意满足 $0\leqslant g\leqslant f$ 的简单函数，如果 h_n 是简单的而且 $0\leqslant h_n\uparrow f$，那么对某个 $c\in[0, \infty]$，$\int h_n \mathrm{d}\mu\uparrow c$. 为证明 $c\geqslant\int g\mathrm{d}\mu$，记 $g=\sum_i a_i 1_{B(i)}$，其中 $B(i)$是不相交的且其并集是 X 的所有元素(因此某个 a_i 可能为 0)．对任意简单函数 $h=\sum_j c_j 1_{A(j)}$，根据命题 4.1.4，

$$h=\sum_i h1_{B(i)}=\sum_{i,j} c_j 1_{B(i)\cap A(j)} \quad 且 \quad \int h\mathrm{d}\mu=\sum_i\int_{B(i)} h\mathrm{d}\mu.$$

因此，只需证明对每个 i，

$$\lim_{n\to\infty}\int_{B(i)} h_n\mathrm{d}\mu\geqslant a_i\mu(B(i)).$$

如果 $a_i=0$，显然成立．否则通过划分 a_i，其中 $a_i>0$，我们可以假定对某个 $E\in\mathcal{S}$，$g=1_E$. 那么给定 $\varepsilon>0$，令 $F_n:=\{x\in E: h_n(x)>1-\varepsilon\}$，有 $F_n\uparrow E$，换句话说，$F_1\subset F_2\subset\cdots$，且 $\bigcup_n F_n=E$. 因此，由可数可加性，$\mu(E)=\mu(F_1)+\sum_{n\geqslant 1}\mu(F_{n+1}/F_n)$，且 $\mu(F_n)\uparrow\mu(E)$. 因此，$c\geqslant(1-\varepsilon)\mu(E)$. 令 $\varepsilon\downarrow 0$，有 $c\geqslant\int g\mathrm{d}\mu$，那么 $c\geqslant\int f\mathrm{d}\mu$. 由于对所有的 n，$\int f_n\mathrm{d}\mu\leqslant\int f\mathrm{d}\mu$，因此，$c=\int f\mathrm{d}\mu$. □

一个 σ-环是一个集类$\mathcal{R}$且$\varnothing\in\mathcal{R}$，使得对任意 $A\in\mathcal{R}$，$B\in\mathcal{R}$，有 $A\setminus B\in\mathcal{R}$，且只要 $A_j\in\mathcal{R}$ $(j=1, 2, \cdots)$，就有 $\bigcup_{j\geqslant 1} A_j\in\mathcal{R}$. 所以 σ-代数是 σ-环．反之，集合 X 的子集构成的 σ-环是 σ-代数，当且仅当 $X\in\mathcal{R}$. 例如，$\mathbb{R}$ 的所有可数子集组成的集合是一个 σ-环，不是 σ-代数．如果 f 是集合 X 上的实值函数，$\mathcal{R}$是由集合 X 的子集构成的 σ-环，那么 f 是$\mathcal{R}$可测的，当且仅当对任意的博雷尔集

$B \subset \mathbb{R}$，$\mathbb{R}$ 不包含 0，$f^{-1}(B) \in \mathcal{R}$.（如果对一般的博雷尔集也有此结论成立，那么$f^{-1}(\mathbb{R}) = X \in \mathcal{R}$意味着$\mathcal{R}$是 σ-代数.）

一个 σ-环称为由$\mathcal{C}$生成的(generated)，当且仅当$\mathcal{R}$是包含$\mathcal{C}$的最小 σ-环，这个结论对 σ-代数同样成立. 下面的事实使验证函数的可测性更为容易.

4.1.6 定理 设$(X, \mathcal{S})$和$(Y, \mathcal{B})$是可测空间，$\mathcal{B}$是由$\mathcal{C}$生成的，那么从 X 映射到 Y 的函数 f 是可测的，当且仅当对所有的 $C \in \mathcal{C}$，$f^{-1}(C) \in \mathcal{S}$. 如果 X 是一个集合，$\mathcal{S}$是 X 的子集组成的 σ-环，$Y = \mathbb{R}$ 且$\mathcal{B}$是$\mathbb{R}$（不包含 0）的博雷尔集组成的 σ-环，那么有同样的结论.

证明 "必要性"显然. 下证充分性：令$\mathcal{D} := \{D \in \mathcal{B}: f^{-1}(D) \in \mathcal{S}\}$，假定$\mathcal{C} \subset \mathcal{D}$，如果对所有的 n，$D_n \in \mathcal{D}$，有$f^{-1}\left(\bigcup_n D_n\right) = \bigcup_n f^{-1}(D_n)$，则 $\bigcup_n D_n \in \mathcal{D}$. 如果 $D \in \mathcal{D}$，$E \in \mathcal{D}$，有 $f^{-1}(E \setminus D) = f^{-1}(E) \setminus f^{-1}(D) \in \mathcal{S}$，则 $E \setminus D \in \mathcal{D}$. 因此，$\mathcal{D}$是 σ-环，$\mathcal{S}$是 σ-代数，且$f^{-1}(Y) = X \in \mathcal{S}$，因此$Y \in \mathcal{D}$. 在另一种情况下，$\mathcal{B} \subset \mathcal{D}$，所以$\mathcal{B} = \mathcal{D}$. □

$\mathbb{R}$ 的子集的一个适当小的集类$\mathcal{C}$是所有半直线(t, ∞)的集合，$\mathbb{R}$ 生成了整个博雷尔 σ-代数. 因此，为了证明实值函数是可测的，只需证对每一个 t，$\{x, f(x) > t\}$是可测的.

设$(X, \mathcal{A})$，$(Y, \mathcal{B})$，$(Z, \mathcal{C})$是可测空间，如果从 X 映射到 Y 的函数 f 是可测的，从 Y 映射到 Z 的函数 g 也是可测的，则因为 $g^{-1}(C) \in \mathcal{B}$，所以对任意的 $C \in \mathcal{C}$，$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C)) \in \mathcal{A}$. 因此从 X 映射到 Z 的复合函数 $g \circ f$ 是可测的（证明基本上与证明连续函数的复合还是连续函数是一样的）.

关于笛卡儿积 $Y \times Z$，令$\mathcal{B} \otimes \mathcal{C}$是由满足 $B \in \mathcal{B}$，$C \in \mathcal{C}$的所有矩形 $B \times C$ 组成的集合生成的 σ-代数，那么称$\mathcal{B} \otimes \mathcal{C}$为 $Y \times Z$ 上的积 σ-代数(product σ-algebra). 对某个从 X 映射到 Y 的函数 f 和从 Y 映射到 Z 的函数 g，从 X 映射到 $Y \times Z$ 的函数 h 是形如 $h(x) = (f(x), g(x))$ 的函数. 根据定理4.1.6，h 是可测的当且仅当 f 和 g 是可测的，对 $B \in \mathcal{B}$，$C \in \mathcal{C}$，考虑矩形 $B \times Z$ 和 $Y \times C$（这样的矩形的集合也生成$\mathcal{B} \otimes \mathcal{C}$）.

118

回想第二可数拓扑有一个可数基（见命题 2.1.4），并且由拓扑可生成一个博雷尔 σ-代数. 当 $X = Y = \mathbb{R}$ 时，下面的事实很有用.

4.1.7 命题 设$(X, \mathcal{T})$，$(Y, \mathcal{U})$是任意两个拓扑空间，对任意的拓扑空间$(Z, \mathcal{V})$，设它的博雷尔 σ-代数是$\mathcal{B}(Z, \mathcal{V})$，那么 $X \times Y$ 上乘积拓扑的 σ-代数$\mathcal{C}$包括乘积 σ-代数$\mathcal{B}(X, \mathcal{T}) \otimes \mathcal{B}(Y, \mathcal{U})$. 如果$(X, \mathcal{T})$和$(Y, \mathcal{U})$是第二可数的，那么 $X \times Y$ 上的两个 σ-代数是相等的.

证明 对任意集合 $A \subset X$，令$\mathcal{U}(A)$是由所有的 $B \in Y$ 组成的集合，使得 $A \times B \in \mathcal{C}$. 如果 A 是开的，那么 $A \times Y \in \mathcal{C}$. 现在 $B \mapsto A \times B$ 保持集合运算，特别地，对任意 $B \subset Y$，$A \times (Y \setminus B) = (A \times Y) \setminus (A \times B)$，且对任意的 $B_n \in Y$，$\bigcup_n (A \times B_n) = A \times \bigcup_n B_n$. 由此可得，$\mathcal{U}(A)$是 Y 的子集的 σ-代数. 它包括$\mathcal{U}$，从而包括$\mathcal{B}(Y, \mathcal{U})$. 这时，对 $B \in \mathcal{B}(Y, \mathcal{U})$，令$\mathcal{T}(B)$是所有 $A \subset X$ 组成的集合，使得 $A \times B \in \mathcal{C}$. 则 $X \in \mathcal{T}(B)$，$\mathcal{T}(B)$是一个 σ-代数. 它包含$\mathcal{T}$，从而包括$\mathcal{B}(X, \mathcal{T})$. 因此博雷尔 σ-代数的乘积 σ-代数包括在乘积的博雷尔 σ-代数$\mathcal{C}$中.

另一方面，假设$(X, \mathcal{T})$和$(Y, \mathcal{U})$是第二可数的. 乘积拓扑有基$\mathcal{W}$，$\mathcal{W}$由所有集合 $A \times B$ 组成，其中 A 是属于$\mathcal{T}$的可数基，B 是属于$\mathcal{U}$的可数基，那么由$\mathcal{W}$生成的 σ-代数是乘积拓扑的博雷尔 σ-代数，显然它包含在乘积 σ-代数中. □

根据命题 2.1.4，$\mathbb{R}$ 上的通常拓扑是第二可数的（或者因为区间(a, b)形成一个基，其中 a，b

是有理数). 因此, 从$\mathbb{R}\times\mathbb{R}$映射到$\mathbb{R}$(或任意拓扑空间)的对博雷尔$\sigma$-代数可测的任意连续函数, 根据命题4.1.7, 它对$\mathbb{R}\times\mathbb{R}$上的乘积$\sigma$-代数也是可测的. 特别地, 从$\mathbb{R}\times\mathbb{R}$映射到$\mathbb{R}$的“和”和“乘”是可测的. 因此, 对任意可测空间$(X,\mathcal{S})$和任意两个X上的实值可测函数f和g、$f+g$和fg是可测的. 令$\mathcal{L}^0(X,\mathcal{S})$为$X$上的所有实值$\mathcal{S}$可测函数组成的集合, 此时由于常值函数是可测的, $\mathcal{L}^0(X,\mathcal{S})$是$\mathbb{R}$上的一个向量空间, 它满足如下的加法和数乘运算: $(f+g)(x):=f(x)+g(x)$和对任意常数c, $(cf)(x):=cf(x)$. 对非负函数, 积分可加.

119

4.1.8 命题 对任意可测空间$(X,\mathcal{S},\mu)$和从X映射到$[0,\infty]$的任意两个可测函数f和g,

$$\int f+g\mathrm{d}\mu=\int f\mathrm{d}\mu+\int g\mathrm{d}\mu.$$

证明 首先, $(f+g)(x)=+\infty$当且仅当$f(x)$或$g(x)$中至少有一个为$+\infty$. 上述情况发生的集合是可测的, $f+g$在它上是可测的. 将$f+g$限制到f和g是有限的集合上, 根据上面的讨论, $f+g$是可测的. 因此, $f+g$是可测的. 根据命题4.1.5, 取简单函数f_n和g_n, 满足$0\leqslant f_n\uparrow f$和$0\leqslant g_n\uparrow g$, 由命题4.1.4知, 对每个n, $\int f_n+g_n\mathrm{d}\mu=\int f_n\mathrm{d}\mu+\int g_n\mathrm{d}\mu$, 再由命题4.1.5, 有

$$\begin{aligned}\int f+g\mathrm{d}\mu&=\lim_{n\to\infty}\int f_n+g_n\mathrm{d}\mu\\&=\lim_{n\to\infty}\left(\int f_n\mathrm{d}\mu+\int g_n\mathrm{d}\mu\right)=\int f\mathrm{d}\mu+\int g\mathrm{d}\mu.\end{aligned}$$

□

通过归纳法, 把命题4.1.8扩张到非负可测函数的任意有限和.

给定任意可测空间$(X,\mathcal{S},\mu)$和从X映射到$[-\infty,\infty]$的可测函数f, 令$f^+:=\max(f,0)$和$f^-:=-\min(f,0)$, 那么f^+和f^-都是非负可测的(正如加和幂次, 取最大和最小均是从$\mathbb{R}\times\mathbb{R}$映射到$\mathbb{R}$的连续函数). 对所有的$x$, 或者$f^+(x)=0$, 或者$f^-(x)=0$, 且$f(x)=f^+(x)-f^-(x)$, 其中这个差总是明确定义的(不是$\infty-\infty$). 我们说积分$\int f\mathrm{d}\mu$是有定义的, 当且仅当$\int f^+\mathrm{d}\mu$和$\int f^-\mathrm{d}\mu$都是有限的. 从而可定义$\int f\mathrm{d}\mu=\int f^+\mathrm{d}\mu-\int f^-\mathrm{d}\mu$, 积分通常和变量一起写, 例如, $\int f(x)\mathrm{d}\mu(x):=\int f\mathrm{d}\mu$. 例如, 如果$f(x):=x^2$, 则$\int x^2\mathrm{d}\mu(x):=\int f\mathrm{d}\mu$. 同样, 如果$\mu$是勒贝格测度$\lambda$, 那么$\mathrm{d}\lambda(x)$通常写为$\mathrm{d}x$.

4.1.9 引理 对任意的可测空间$(X,\mathcal{S},\mu)$和从X映射到$[-\infty,\infty]$的两个可测函数$f\leqslant g$, 只可能有下面的情况出现:

(a) $\int f\mathrm{d}\mu\leqslant\int g\mathrm{d}\mu$(这两个积分都有定义).

(b) $\int f\mathrm{d}\mu$无定义, $\int g\mathrm{d}\mu=+\infty$.

(c) $\int f\mathrm{d}\mu=-\infty$, $\int g\mathrm{d}\mu$无定义.

(d) 两个积分都无定义.

证明 如果$f\geqslant 0$, 直接从定义必可得结论(a). 一般情况下, 有$f^+\leqslant g^+$和$f^-\geqslant g^-$, 从而如果两个积分都有定义, 那么(a)成立. 如果$\int f\mathrm{d}\mu$无定义, 那么$+\infty=\int f^+\mathrm{d}\mu\leqslant\int g^+\mathrm{d}\mu$, 所以$\int g^+\mathrm{d}\mu=+\infty$且$\int g\mathrm{d}\mu$无定义或$+\infty$. 类似地. 如果$\int g\mathrm{d}\mu$无定义, $-\infty=\int -g^-\mathrm{d}\mu\geqslant\int -f^-\mathrm{d}\mu$, 所以$\int f\mathrm{d}\mu$无定

[120] 义或为 $-\infty$. □

从 X 映射到 $\mathbb{R}$ 且满足 $\int |f| \mathrm{d}\mu < +\infty$ 的可测函数称为是可积的(integrable). 所有关于 μ 可积的函数组成的集记为 $\mathcal{L}^1(X, \mathcal{S}, \mu)$, 这个集合也简记为 $\mathcal{L}^1(\mu)$ 或 $\mathcal{L}^1$.

4.1.10 定理 在 $\mathcal{L}^1(X, \mathcal{S}, \mu)$ 中, $f \mapsto \int f \mathrm{d}\mu$ 是线性的, 即对任意的 $f, g \in \mathcal{L}^1(X, \mathcal{S}, \mu)$ 和 $c \in \mathbb{R}$, 有 $\int cf \mathrm{d}\mu = c\int f \mathrm{d}\mu$ 和 $\int f+g \mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu$. 当 $f \in \mathcal{L}^1(X, \mathcal{S}, \mu)$ 且 g 为任意非负可测函数时, 后者也成立.

证明 $f+g$ 和 cf 是可测的(参见命题 4.1.7). 根据定义, 如果 $c=-1$, 那么有 $\int cf \mathrm{d}\mu = c\int f \mathrm{d}\mu$, 对 $c \geqslant 0$, 由命题 4.1.5, 也有 $\int cf \mathrm{d}\mu = c\int f \mathrm{d}\mu$, 所以对所有的 $c \in \mathbb{R}$, 有 $\int cf \mathrm{d}\mu = c\int f \mathrm{d}\mu$.

如果 $f, g \in \mathcal{L}^1$, 那么对 $h := f+g$, 有 $f^- + g^- + h^+ = f^+ + g^+ + h^-$. 因为 $h^- = 0$, 其中 $h \geqslant 0$, 从而 $h^+ \leqslant f^+ + g^+$, 所以根据命题 4.1.8 和引理 4.1.9, 有 $\int h^+ \mathrm{d}\mu < +\infty$. 类似地, $\int h^- \mathrm{d}\mu < +\infty$, 所以 $h \in \mathcal{L}^1(X, \mathcal{S}, \mu)$. 由命题 4.1.8 和定义, 有

$$\int h \mathrm{d}\mu = \int h^+ \mathrm{d}\mu - \int h^- \mathrm{d}\mu = \int f^+ \mathrm{d}\mu + \int g^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu - \int g^- \mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu.$$

相反, 如果 $g \geqslant 0$ 且 g 是可测的, 剩下的情况是 $\int g \mathrm{d}\mu = +\infty$. 这时记 $g \leqslant (f+g)^+ + f^-$ ($f \geqslant 0$ 时这是显然的; 对 $f<0$, $g = (f+g) - f \leqslant (f+g)^+ + f^-$). 根据命题 4.1.8 和引理 4.1.9, 有 $+\infty = \int g \mathrm{d}\mu \leqslant \int (f+g)^+ \mathrm{d}\mu + \int f^- \mathrm{d}\mu$. 由于 $\int f^- \mathrm{d}\mu$ 是有限的, 这给出 $\int (f+g)^+ \mathrm{d}\mu = +\infty$. $(f+g)^- \leqslant f^-$ 意味着 $\int (f+g)^- \mathrm{d}\mu$ 是有限的, 所以 $\int f+g \mathrm{d}\mu = +\infty = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu$. □

函数(尤其当它们不是实值的时候)可以称为是变换或映射. 令 $(X, \mathcal{S}, \mu)$ 是测度空间, $(Y, \mathcal{B})$ 是可测空间, T 是从 X 映射到 Y 的可测变换. 那么对所有的 $A \in \mathcal{B}$, 令 $(\mu \circ T^{-1})(A) := \mu(T^{-1}(A))$, 由于 $A \mapsto T^{-1}(A)$ 保持所有的集合运算, 比如可数并和不相交, $\mu \circ T^{-1}$ 是一个可数可加测度. 如果 μ 是有限的, 但不一定是 σ-有限的(令 T 是常值映射), 那么 $\mu \circ T^{-1}$ 是有限的. 这里 $\mu \circ T^{-1}$ 称为 μ 在 T 下的像测度. 例如, 如果 μ 是勒贝格测度, $T(x) = 2x$, 那么 $\mu \circ T^{-1} = \mu/2$. 关于测度的积分和关于其像的积分是通过一个简单的"变量替换定理"联系起来的.

4.1.11 定理 设 f 是从 Y 映射到 $[-\infty, \infty]$ 的任意可测函数, 如果两个积分之一有定义(可能无

[121] 限), 则有 $\int f \mathrm{d}(\mu \circ T^{-1}) = \int f \circ T \mathrm{d}\mu$.

证明 如果对某个 A 和 $c \geqslant 0$, 有 $f = c1_A$, 则结论显然成立. 因此根据命题 4.1.8, 对任意的非负简单函数, 结论都成立. 由命题 4.1.5 可得, 对任意的可测函数 $f \geqslant 0$, 结论也成立. 这时, 取 f^+ 和 f^-, 因为 $(f \circ T)^+ = f^+ \circ T$ 和 $(f \circ T)^- = f^- \circ T$, 所以由 $\int f \mathrm{d}\mu$ 的定义可知, 对任意的可测函数 f, 结论也成立. □

习题

1. 对可测空间$(X, \mathcal{S}, \mu)$，设f为简单函数，g为μ-简单函数，证明：fg是μ-简单的．
2. 在$[0, \infty)$上，$f \equiv x$，在对f构造f_n中(命题4.1.5)，f_4的最大值是多少？在它的值域中f_4有多少不同的值？
3. 对$f, g \in \mathcal{L}^1(X, \mathcal{S}, \mu)$，设$d(f, g) = \int |f-g| \mathrm{d}\mu$，证明：$d$和$\mathcal{L}^1$上的伪测度.
4. 证明：μ-简单函数构成的集合在$\mathcal{L}^1$上关于d是稠密的.
5. 在非负整数集合$\mathbb{N}$上，设c是计数测度：当E有限时，$c(E) = \mathrm{card}E$，当E无限时，$c(E) = +\infty$对E无限，证明：对$f := \mathbb{N} \mapsto \mathbb{R}$，$f \in \mathcal{L}^1(\mathbb{N}, 2^{\mathbb{N}}, c)$当且仅当$\sum_n |f(n)| < +\infty$，从而$\int f\mathrm{d}\mu = \sum_n |f(n)|$.
6. 对任意的集合A，$f[A] := \{f(x) : x \in A\}$，给定两个集合$B$和$C$，令$D := f[B] \cup f[C]$，$E := f[B \cup C]$，$F := f[B] \cap f[C]$以及$G := f[B \cap C]$，证明：对所有的$B$和$C$，下面的包含关系成立，或通过举反例证明不成立.
   (a)$D \subset E$；(b)$E \subset D$；(c)$F \subset G$；(d)$G \subset F$.
7. 如果$(X, \mathcal{S}, \mu)$是一个测度空间，f是X上的非负可测函数，对任意的集合$A \in \mathcal{S}$，令$(f\mu)(A) := \int_A f\mathrm{d}\mu$.
   (a)证明：$f\mu$是一个测度．
   (b)如果T是可测的且对可测空间$(Y, \mathcal{A})$，T是从X映射到Y的一一对应，逆测度为T^{-1}，证明：$(f\mu) \circ T^{-1} = (f \circ T^{-1})(\mu \circ T^{-1})$.
8. 设f是$\mathbb{R}^2$上的由$f := \sum_{j=1}^{n} j 1_{(j,j+2] \times (j,j+2]}$定义的简单函数，求由矩形$(j, j+2] \times (j, j+2]$ $(j=1, 2, \cdots, n)$生成的代数的原子，再把f表示成这些原子的常数次幂示性函数的和．
9. 设$\mathcal{R}$是集合X的子集组成的σ-环，$\mathcal{S}$是由$\mathcal{R}$生成的σ-代数，回顾(3.3节习题8)或证明$\mathcal{S}$包含$\mathcal{R}$中的所有集合和$\mathcal{R}$中所有集合的补集．
   (a)μ从$\mathcal{R}$映射到$[0, \infty]$是可数可加的，对任意的集合$C \subset X$，令$\mu_*(C) := \sup\{\mu(B) : B \subset C, B \in \mathcal{R}\}$(内测度)，证明：$\mu_*$限制到$\mathcal{S}$是一个测度，而且在$\mathcal{R}$上等于$\mu$.
   (b)μ在$\mathcal{S}$上的扩张测度是唯一的，当且仅当对任意的集合$A \in \mathcal{R}$，或者$\mathcal{S} = \mathcal{R}$，或者$\mu_*(X \setminus A) = +\infty$.

122

10. 设$(S, \mathcal{T})$是第二可数的拓扑空间，(Y, d)是任意度量空间，证明：乘积$S \times Y$上的博雷尔σ-代数是S和Y中的博雷尔σ-代数的乘积σ-代数．[提示：这是对命题4.1.7的改进．设V是$S \times Y$中的任意开集，$\{U_m\}_{m \geq 1}$是$\mathcal{T}$的可数基，对每个m和$r > 0$，令
$$V_{mr} := \{y \in Y: \text{对某个}\ \delta > 0, U_m \times B(y, r+\delta) \subset V\},$$
其中$B(y, t) := \{v \in Y : d(y, v) < t\}$，证明在$Y$和$V = \bigcup_{m,n} U_m \times V_{m,1/n}$中，每个$V_{mr}$都是开的．
11. 证明：对某些拓扑空间$(X, \mathcal{S})$和$(Y, \mathcal{T})$，在具有乘积拓扑的$X \times Y$中存在闭集，其中乘积拓扑不在任意乘积σ-代数$\mathcal{A} \otimes \mathcal{B}$中，例如，如果$\mathcal{A}$和$\mathcal{B}$是给定拓扑的博雷尔$\sigma$-代数．[提示：令$X = Y$是势比$c$大的集合，例如，$I := [0, 1]$的所有子集组成的集合$2^I$(定理1.4.2)，设$D$是对角集合$\{(x, x) : x \in X\}$，证明对每个$C \in \mathcal{A} \otimes \mathcal{B}$，存在序列$\{A_n\} \subset \mathcal{A}$和$\{B_n\} \subset \mathcal{B}$，使得$C$在由$\{A_n \times B_n\}_{n \geq 1}$生成的$\sigma$-代数中．对每个$n$，令$x =_n u$意味着$x \in A_n$当且仅当$u \in A_n$，定义一个关系$x \equiv u$当且仅当对所有的$n$，$x =_n u$. 证明这是一个至多有$c$个不同等价类的等价关系，而且对任意的$x$，$y$和$u$，如果$x \equiv u$，那么$(x, y) \in C$当且仅当$(u, y) \in C$. 因为$C = D$且$y = x$，矛盾．]

*4.2　可测性

设$(Y, \mathcal{T})$是一个拓扑空间，它的博雷尔集$\mathcal{B} := \mathcal{B}(Y) := \mathcal{B}(Y, \mathcal{T})$的$\sigma$-代数是由$\mathcal{T}$生成的．如果$(X, \mathcal{S})$是一个可测空间，从$X$映射到$Y$的函数$f$称为是可测的，当且仅当对所有的$B \in \mathcal{B}$，有

$f^{-1}(B) \in \mathcal{S}$(除非指定 Y 的另一个 σ-代数). 如果 X 是实直线$\mathbb{R}$，它的博雷尔集的 σ-代数是$\mathcal{B}$且勒贝格可测集的 σ-代数是$\mathcal{L}$，f称为是博雷尔可测的当且仅当在($\mathbb{R}$，$\mathcal{B}$)上，它是可测的；f称为勒贝格可测的当且仅当在($\mathbb{R}$，$\mathcal{L}$)上，它是可测的. 注意到在这两种情况下，博雷尔 σ-代数在值域空间中都适用. 事实上，本节的主题是可测性在取$\mathbb{R}$或任意完备可分度量空间作为值域空间并且在其上使用博雷尔 σ-代数时是很好用的. 奇异型——它与其他的 σ-代数或其他的值域空间可以导致错误结论——在这个阶段不重要. 本节的命题 4.2.3 给出了一个例子. 奇异型的进一步讨论参见附录 E. 它说明了为什么本书中关于局部紧空间的介绍比其他以往的课本中要少得多. 本节的其余部分可以略读作为以后的参考.

123

下面的事实说明了为什么$\mathbb{R}$上的勒贝格 σ-代数作为值域可能太大.

4.2.1 命题 *存在从 $I:=[0, 1]$映射到其自身的连续非递减函数 f 和一个勒贝格可测集 L，使得 $f^{-1}(L)$不是勒贝格可测的(假设选择公理成立).*

证明 与康托尔集 C(正如命题3.4.1 的证明中那样)相联系的是康托尔函数 g，定义如下：从 I 映射到其自身，g 是非递减的连续函数，满足在$[1/3, 2/3]$上，$g=1/2$，在$[1/9, 2/9]$上，$g=1/4$，在$[7/9, 8/9]$上，$g=3/4$，在$[1/27, 2/27]$上，$g=1/8$，如此下去(参见图 4-2).

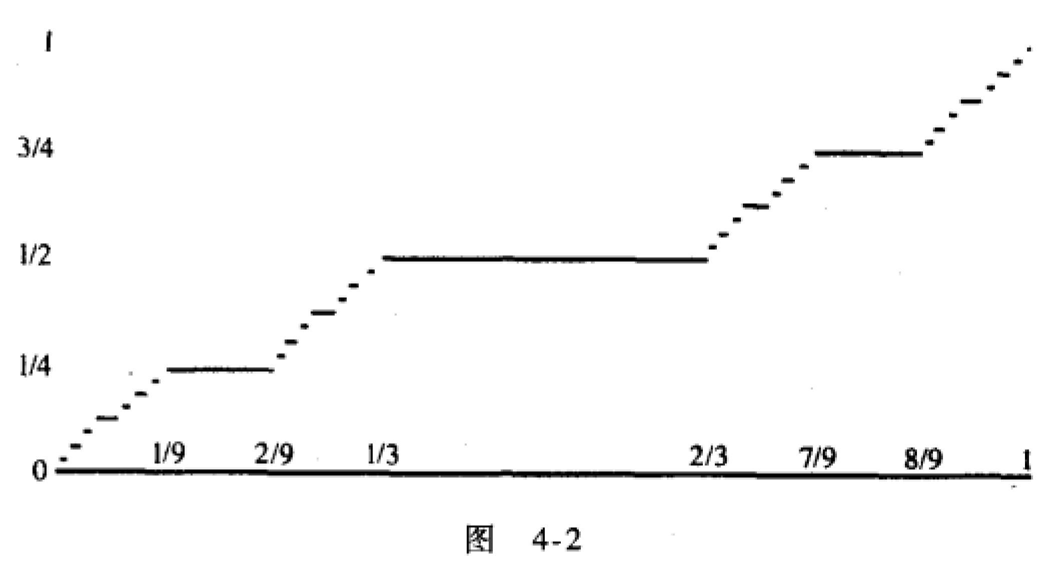


图 4-2

这里 g 可以表达成如下形式. 每个 $x\in[0, 1]$有一个三重扩张 $x = \sum_{n\geq 1} x_n/3^n$，其中对所有的 n，$x_n=0$，1 或 2. 当 I 中所有其他数字仅有一个时，数字 $m/3^n(m\in\mathbb{N})$，$0<m<3^n$ 有两个这样的扩张. 由于 C 是所有 x 组成的集合，x 有一个扩张满足对所有的 n，$x_n\neq 1$. 因此对 $x\notin C$，令$j(x)$是最小的j，使得 $x_j=1$. 如果 $x\in C$，令 $j(x) = +\infty$，则

$$g(x) = 1/2^{j(x)} + \sum_{i=1}^{j(x)-1} x_i/2^{i+1}, \quad 0\leqslant x\leqslant 1.$$

由此可以说明 g 是非递减连续的(Halmos，1950，p. 83，给出了提示)，但是这些性质从图中看很明显. 现在 g 在双值有理数$\{m/2^n: m=1, \cdots, 2^n-1, n=1, 2, \cdots\}$上取 $I\setminus C$，因为g 在 I 上取 I，g 必须在 I 上取 C(在 C 的补集中的每个"中间第三"个区间上的取值取为 C 中的终点值). 对 $0\leqslant x\leqslant 1$，令 $h(x):=(g(x)+x)/2$，那么 h 是从 I 映上到自身的连续严格递增函数，即对 $t<u$，$h(t)<h(u)$. $I\setminus C$ 中每个开的中间第三个区间都是在区间一半长度上取. 因此，在满足 $\lambda(U)=1/2$ 的开集 U 上取 $I\setminus C$(由命题3.4.1 知，$\lambda(C)=0$，所以 $\lambda(I\setminus C)=1$). 令$f=h^{-1}$，则f是从 I 映上到自身的连续严格递增函数，满足$f^{-1}(C)=h[C]=I\setminus U:=F$，那么 $\lambda(F)=1/2$ 且 F 的任意子集是形如$f^{-1}(L)$的，其中 $L\subset C$，所以 L 是勒贝格可测的且满足 $\lambda(L)=0$. 令 E 是 I 中的非可测子集，满足

124

$\lambda^*(E)=\lambda^*(I\setminus E)=1$，根据命题3.4.4，有 E 和 $I\setminus E$ 都不包含任何勒贝格可测集合 A，使得 $\lambda(A)>0$，因此，$E\cap F$ 和 $F\setminus E$ 都不包含这样的集合．F 是 $E\cap F$ 的一个可测覆盖(参见3.3节)，如果 F 不是而 G 是，$F\setminus E$ 也许包括一个有正测度的可测集．所以 $\lambda^*(E\cap F)=1/2$．类似地，$\lambda^*(F\setminus E)=1/2$，所以 $\lambda^*(E\cap F)+\lambda^*(F\setminus E)=1\neq\lambda(F)=1/2$，且 $E\cap F$ 不是勒贝格可测的．□

下面的两个事实是与可测函数序列的极限有关．为看到这里有些结论不是平凡的，设 f_n 是某一集合 X 上的实值函数序列，满足对所有的 $x\in X$，$f_n(x)$ 收敛到 $f(x)$，设 U 是开区间 (a,b)，注意到，对所有的 n，可能有 $x\in f_n^{-1}(U)$，但是 $x\notin f^{-1}(U)$(如果 $f(x)=a$)，因此 $f^{-1}(U)$ 不能用集合 $f_n^{-1}(U)$ 表示．

4.2.2 定理 设 $(X,\mathcal{S})$ 是一个可测空间，(Y,d) 是一个度量空间，f_n 是从 X 映射到 Y 的可测函数，使得对所有的 $x\in X$，在 Y 中有 $f_n(x)\to f(x)$，那么 f 是可测的．

证明 根据定理4.1.6，只需证明对 Y 中任意的开集 U，有 $f^{-1}(U)\in\mathcal{S}$．令 $F_m:=\{y\in U: B(y,1/m)\in U\}$，其中 $B(y,r):=\{v: d(v,y)<r\}$．那么 F_m 是闭的：如果对所有的 j，$y_j\in F_m$，有 $y_j\to y$ 且 $d(y,v)<1/m$，则对足够大的 j，$d(y_j,v)<1/m$，所以 $v\in U$．现在 $f(x)\in U$ 当且仅当对某个 m，$f(x)\in F_m$，从而对足够大的 n，$d(f_n(x),f(x))<1/(2m)$，即对足够大的 n，$f_n(x)\in F_{2m}$．反之对足够大的 n，$f_n(x)\in F_m$，则有 $f(x)\in F_m\subset U$，因此

$$f^{-1}(U)=\bigcup_m\bigcup_k\bigcap_{n\geq k}f_n^{-1}(F_m)\in\mathcal{S}.$$ □

125

现在设 $I=[0,1]$，具有通常拓扑，那么根据吉洪诺夫定理2.2.8知，具有乘积拓扑的 I^I 是紧豪斯多夫空间．这个空间有很多很好的性质，但是根据下面的事实，定理4.2.2不能推广到这些空间上(作为值域空间)．它的证明假设了选择公理(通常，特别是在处理如 I^I 这样的空间时)．

4.2.3 命题 存在从 I 映射到 I^I 的连续(因此博雷尔可测)函数序列 f_n，f_n 满足对所有的 $x\in I$，$f_n(x)\in I^I$，$f_n(x)\to f(x)\in I^I$，但是 f 不是勒贝格可测的：存在开集 $W\subset I^I$，使得 $f^{-1}(W)$ 在 I 中不是勒贝格可测集．

证明 对 $x,y\in I$，令 $f_n(x)(y):=\max(0,1-n|x-y|)$，验证 f_n 是连续的，只需考虑乘积拓扑的一般子基．对任意的开集 $V\subset I$ 和 $y\in I$，正如所期望的 $\{x\in I: f_n(x)(y)\in V\}$ 在 I 中是开的．令

$$\text{当 } x=y \text{ 时，} f(x)(y):=1_{x=y}=1\text{，否则为 }0.$$

那么对所有的 $x,y\in I$，当 $n\to\infty$ 时，$f_n(x)(y)\to f(x)(y)$．从而在 I^I 中，对所有的 $x\in I$，有 $f_n(x)\to f(x)$．现在令 E 是 I 中的任意子集，$W:=\{g\in I^I$：对某个 $y\in E$，$g(y)>1/2\}$，那么 W 在 I^I 中是开的且 $f^{-1}(W)=E$，其中 E 不一定是勒贝格可测的(定理3.4.4)．□

如果 $(X,\mathcal{S})$ 是一个可测空间，$A\subset X$，令 $\mathcal{S}_A:=\{B\cap A: B\in\mathcal{S}\}$．那么 $\mathcal{S}_A$ 是 A 的子集的一个 σ-代数，且 $\mathcal{S}_A$ 称为(在 A 上 $\mathcal{S}$ 的)相对 σ-代数．下面给出一个经常使用的简单易懂的引理．

4.2.4 引理 设 $(X,\mathcal{S})$ 和 $(Y,\mathcal{B})$ 是可测空间，E_n 是 $\mathcal{S}$ 中不相交的集合且 $\bigcup_n E_n=X$，对每个 $n=1,2,\cdots$．令 f_n 从 E_n 映射到 Y 是可测的且有相对 σ-代数．对所有的 $x\in E_n$，定义 f 为 $f(x)=f_n(x)$，则 f 是可测的．

证明 因为每个 $E_n\in\mathcal{S}$，对任意的 $B\in\mathcal{B}$，$f_n^{-1}(B)\in\mathcal{S}$ 等价于对某个 $A_n\in\mathcal{S}$，$f_n^{-1}(B)=A_n\cap E_n$．因此，

$$f^{-1}(B)=\bigcup_n f_n^{-1}(B)\in\mathcal{S}.$$ □

如果f是X上的任意可测函数，那么限制到A上的f是$\mathcal{S}_A$可测的．同理，限制到一个子集上的任意连续函数对相对拓扑是连续的．但反之，一个对相对拓扑连续的函数并不总是可以扩张为一个更大的集合上的连续函数．例如，在$(0, 1)$上，$1/x$不能扩张为$[0, 1)$上的连续实值函数，有界函数$\sin(1/x)$也不可以．根据扩张定理2.6.4，（从正规空间X的一个闭子集映射到$\mathbb{R}$的连续函数，比如一个度量空间，总是可以扩张到所有的X上）．一般地，定义在可测集A上的可测函数f可以扩张为在X上的可测函数g，例如，令g在$X \setminus A$上有固定的值．实值可测函数总是可以扩张的，即使A不可测，这一问题不是很明显但却是正确的．

126

4.2.5 定理 *设$(X, \mathcal{S})$是任意的可测空间，A是X的任意子集（不一定在$\mathcal{S}$中），f是A上的$\mathcal{S}_A$可测实值函数，那么f可以扩张为X上的所有实值函数，且是$\mathcal{S}_A$可测的．*

证明 令$\mathcal{G}$是A上所有$\mathcal{S}_A$可测实值函数组成的集合，A有$\mathcal{S}$可测的扩张，那么显然$\mathcal{G}$是一个向量空间，且对每个$S \in \mathcal{S}$，$1_{A\cap S}$有扩张1_S，所以$\mathcal{G}$包含了对$\mathcal{S}_A$可测的所有简单函数．为了证明$f \in \mathcal{G}$，假设$f \geqslant 0$，因为$f^+ \in \mathcal{G}$和$f^- \in \mathcal{G}$，从而$f \in \mathcal{G}$．令f_n是满足$0 \leqslant f_n \uparrow f$的简单函数（对$\mathcal{S}_A$），根据命题4.1.5，令$g_n$扩张$f_n$．令$g(x) := \lim_{n\to\infty} g_n(x)$，该极限存在（且有限）．否则令$g(x) = 0$．显然，$g$扩张$f$．使得$g_n(x)$收敛或等价于$g_n(x)$是柯西序列的$x$构成的集合是$G := \bigcap_{k\geqslant 1} \bigcup_{n\geqslant 1} \bigcap_{m\geqslant n} \left\{x: |g_m(x) - g_n(x)| < \frac{1}{k}\right\}$．因此$G \in \mathcal{S}$．令在$G$上，$h_n := g_n$，在$X \setminus G$上，$h_n := 0$，那么根据引理4.2.4，每个$h_n$是可测的，且对所有的$x$，$h_n(x) \to g(x)$．从而根据定理4.2.2，$g$是$\mathcal{S}$可测的．□

由定理4.2.2和下面的命题知，定理4.2.5中的值域空间$\mathbb{R}$可以由具有博雷尔σ-代数的任意完备可分度量空间代替．

4.2.6 命题 *对任意的可分度量空间(S, d)，由S映射到自身的恒等函数是由从S映射到自身的博雷尔可测函数序列的逐点极限，其中每个f_n都有有限的值域且对所有的x，当$n\to\infty$时，$f_n(x) \to x$．*

证明 令$\{x_n\}$是S中的可数稠密集．对每个$n = 1, 2, \cdots$，令$f_n(x)$是$x_1, x_2, \cdots, x_n$中最接近x的点，或如果有两个或更多个同等接近的点就取有最小指标的那个点．那么f_n的值域包含在$\{x_1, \cdots, x_n\}$中，且对每个$j \leqslant n$，

127

$$f_n^{-1}(\{x_j\}) = \bigcap_{i<j} \{x: d(x, x_i) > d(x, x_j)\} \cap \bigcap_{j\leqslant i\leqslant n} \{x: d(x, x_i) \geqslant d(x, x_j)\}.$$

后者是一个开集和一个闭集的交，因此是一个博雷尔集，所以f_n是可测的．显然，f_n逐点收敛到恒等函数．□

给定一个可测空间$(X, \mathcal{S})$，X上的函数g称为是简单的，当且仅当它的值域Y是有限的且对每个$y \in Y$，$g^{-1}(\{y\}) \in \mathcal{S}$．

4.2.7 推论 *对任意的可测空间$(U, \mathcal{S})$，$X \subset U$，非空可分度量空间(S, d)和从X映射到S的$\mathcal{S}_X$可测函数，存在从X映射到S的简单函数g_n，使得对所有的x，$g_n(x) \to g(x)$．如果S是完备的，则g可作为一个$\mathcal{S}$可测函数扩张到所有的U上．*

证明 令$g_n := f_n \circ g$，f_n是命题4.2.6中定义的，则对所有的x，g_n是简单的且对所有的x，$g_n(x) \to g(x)$．这里每个g_n都可定义在所有的U上．如果S是完备的，剩下的证明参见定理4.2.5（S中的一个任意点代替0）．□

现在设$(Y, \mathcal{B})$是一个可测空间，X是任意集合，T是一个从X映射到Y的函数．令$T^{-1}[\mathcal{B}] := \{T^{-1}(B): B \in \mathcal{B}\}$，那么$T^{-1}[\mathcal{B}]$是$X$的子集的一个$\sigma$-代数．

4.2.8 定理 *给定一个集合X，一个可测空间$(Y, \mathcal{B})$和一个从X映射到Y的函数T，X上的一个实*

值函数 f 在 X 上是 $T^{-1}[\mathcal{B}]$ 可测的，当且仅当对 Y 上的某个 $\mathcal{B}$ 可测函数，有 $f=g\circ T$.

证明　"充分性"显然．"必要性"，如果 f 是 $T^{-1}[\mathcal{B}]$ 可测的，那么只要 $T(u)=T(v)$，就有 $f(u)=f(v)$，如果不成立，令 B 是 $\mathbb{R}$ 上满足 $f(u)\in B$ 和 $f(v)\notin B$ 的一个博雷尔集．对某个满足 $T(u)\in C$ 但 $T(v)\notin C$ 的 $C\in\mathcal{B}$，有 $f^{-1}(B)=T^{-1}(C)$，由此得出矛盾．因此，对某个从 $D:=\operatorname{ran}T$ 映射到 $\mathbb{R}$ 的函数 g，有 $f=g\circ T$. 对任意的博雷尔集 $S\subset\mathbb{R}$，对某个 $F\in\mathcal{B}$，有 $T^{-1}(g^{-1}(S))=f^{-1}(S)=T^{-1}(F)$，所以 $F\cap D=g^{-1}(S)$ 且 g 是 $\mathcal{B}_D$ 可测的．根据定理4.2.5，g 有一个 $\mathcal{B}$ 测度可以扩张到所有的 Y 上．□

习题

[128]

1. 设 $(X,\mathcal{S})$ 是一个可测空间，E_n 是可测集，不一定不相交，其并为 X. 假设对每个 n，f_n 是 E_n 上的实值可测函数，对任意 $x\in E_m\cap E_n$，以及任意的 m 和 n，有 $f_m(x)=f_n(x)$，对任意的 $x\in E_n$，及任意的 n，令 $f(x):=f_n(x)$. 证明：f 是可测的.
2. 设 $(X,\mathcal{S})$ 是一个可测空间，f_n 是从 X 映射到 $[-\infty,\infty]$ 的任意可测函数序列，证明：
   (a) 定义 $f(x):=\sup_n f_n(x)$，则 f 是一个可测函数.
   (b) 定义 $\lim\limits_{n\to\infty}\inf f_n:=\sup\limits_n\inf\limits_{m\geqslant n}f_n$，$g(x):=\lim\limits_{n\to\infty}\sup f_n(x):=\inf\limits_m\sup\limits_{m\geqslant n}f_m(x)$，则可测函数 g 和 $\lim\limits_{n\to\infty}\inf f_n:=\sup\limits_n\inf\limits_{m\geqslant n}f_n$ 都为可测函数.
3. 证明或举反例：设 f 是从 $[0,1]$ 映射到自身的连续严格递增函数，使得几乎对所有的 x（勒贝格测度），其导数 $f'(x)$ 存在（"严格递增"意味着对 $0\leqslant x<y\leqslant 1$，$f(x)<f(y)$），那么对所有的 x，$\int f'(t)\,\mathrm{d}t=f(x)-f(0)$.
   [提示：参见命题4.2.1.]
4. 设 $f(x):=1_{\{x\}}$，从而 f 定义了一个从 I 映射到 I^I 的函数，如同在命题4.2.3中定义的．
   (a) 证明：f 的值域是 I^I 中的一个博雷尔集．
   (b) 证明：在 $I\times I^I$（具有乘积拓扑）中 f 的图像是一个博雷尔集．
5. 证明或举反例：习题4中的函数 f 是具有有限值域的函数序列的极限．
6. 在定理4.2.8中，设 $X=\mathbb{R}$，Y 是 $\mathbb{R}^2$ 中的单位圆：$Y:=\{(x,y):x^2+y^2=1\}$，$\mathcal{B}$ 是 Y 上的博雷尔 σ-代数，对所有的 $u\in\mathbb{R}$，令 $T(u):=(\cos u,\sin u)$，求下面哪些 $\mathbb{R}$ 上的函数 f 是 $T^{-1}[\mathcal{B}]$ 可测的，且对这些函数，像在定理4.2.8中那样，求出函数 g.
   (a) $f(t)=\cos(2t)$；(b) $f(t)=\sin(t/2)$；(c) $f(t)=\sin^2(t/2)$.
7. 对康托尔函数 g，定义见命题4.2.1的证明，对 $k=0,1,2,3,4,5,6,7$ 和 8，求 $g(k/8)$ 的值．
8. 证明：$\mathbb{R}$ 中的博雷尔集类与 $\mathbb{R}$ 有相同的势 c.［提示：易证，至少有一个 c 博雷尔集．然后取一个不可数良序集 $(J,<)$，使得对每个 $j\in J$，$\{i\in J:i<j\}$ 是可数的．令 α 是 J 中最小的元素，递归地定义 $\mathbb{R}$ 中的博雷尔集的集类 $\mathcal{B}_j$：令 $\mathcal{B}_\alpha$ 为由具有有理终点的所有开区间构成的集类．对任意的 $\beta\in J$，如果 β' 是下一个最大的元素，递归地令 $\mathcal{B}_{\beta'}$ 为 $\mathcal{B}_\beta$ 中集合的所有余集和可数并构成的集类．如果 $\gamma\in J$ 没有直接前趋（对所有的 β，$\gamma\neq\beta$），$\gamma>\alpha$，对所有 $\beta<\gamma$，令 β_γ 是 $\mathcal{B}_\beta$ 的并．证明对 $\beta\in J$，所有 $\mathcal{B}_\beta$ 的并是所有博雷尔集的集类，且它的势为 c.（参见1.4节的习题5.）］

[129]

9. 设 f 是从 X 映上到 S 的可测函数，其中 $(X,\mathcal{A})$ 是一个可测空间，(S,e) 是一个有博雷尔 σ-代数的度量空间．令 T 是 S 的子集，S 有离散相对拓扑（T 的所有子集在 T 中是开的），证明：存在从 X 映上到 T 的可测函数 g.［提示：对充分接近 $t\in T$ 的 $f(x)$，令 $g(x)=t$；否则对固定的 $t_0\in T$，令 $g(x)=t_0$.］
10. 设 f 是从可分度量空间 X 映上到具有度量 e 的度量空间 S 的一个博雷尔可测函数，证明：(S,e) 是可分的．［提示：同习题8，X 至多有 c 个博雷尔集．如果 S 不是可分的，那么证明对某个 $\varepsilon>0$，存在 S 的一个不可

数子集 T，满足在 T 中，对所有的 $y\neq z$，$d(y,z)>\varepsilon$. 利用习题 9 得到一个从 X 映上到 T 的可测函数．所有 $g^{-1}(A)(A\subset T)$ 是 X 中的博雷尔集．]

4.3 积分收敛定理

在本节中，令 $(X,\mathcal{S},\mu)$ 是一个测度空间．关于 $x\in X$ 称为是几乎处处(almost everywhere)成立，或 a. e.，当且仅当对某个 A，对所有的 $x\notin A$，$\mu(A)=0$，它成立．(就像在 3.3 节中那样，所有使此陈述成立的 x 构成的集合将对 μ 的完备化测度是可测的，但是这个集合不一定必须在 $\mathcal{S}$ 中) 这一陈述也可说成是对几乎所有的 x 成立．例如，对勒贝格测度，$1_{[a,b]}=1_{(a,b)}$ a. e..

4.3.1 命题 如果 f 和 g 是从 X 映射到 $[-\infty,\infty]$ 的两个可测函数，且 $f(x)=g(x)$ a. e.，那么 $\int f\mathrm{d}\mu$ 是有定义的，当且仅当 $\int g\mathrm{d}\mu$ 有定义．当二者均有定义时，两个积分相等．

证明 在 $X\setminus A$ 上，令 $f=g$，其中 $\mu(A)=0$，我们证明 $\int h\mathrm{d}\mu=\int_{X\setminus A}h\mathrm{d}\mu$，其中 h 是任意的可测函数，且此积分有定义且相等的情况下，等式成立当且仅当两者之一是有定义的．如果 h 是 S 中集合的示性函数，结论显然成立；如果 h 是任意的非负简单函数，结论也成立；于是根据命题 4.1.5，如果 h 是任意的非负可测函数，结论也成立．从而对一般的函数 h，根据积分的定义，令 $h=f$ 和 $h=g$ 完成证明． □

设 f 为定义在满足 $\mu(X\setminus B)=0$ 的集合 $B\in\mathcal{S}$ 上的一个函数，其中 f 在 $[-\infty,\infty]$ 中取值且是 $\mathcal{S}_B$ 可测的．那么 f 可扩张为 X 上的一个 $\mathcal{S}$ 可测函数(例如，在 $X\setminus B$ 上，令 $f=0$)．对 f 到 X 上的任意两个扩张 g 和 h，$g=h$ a. e.. 因此，如果这个有定义，则可以把 $\int f\mathrm{d}\mu$ 定义为 $\int g\mathrm{d}\mu$. 从而根据命题 4.3.1，$\int f\mathrm{d}\mu$ 是有明确定义的．如果对 $n=1,2,\cdots$，$f_n=g_n$ a. e.，那么 $\mu^*\left(\bigcup_n\{x: f_n(x)\neq g_n(x)\}\right)=0$. 在这个集合之外，对所有的 n，$f_n=g_n$. 因此，在关于积分的定理中，甚至对于下面的函数序列，只需假设几乎处处成立．

[130]

下面的三个定理是分析中最重要和最常用的定理．

4.3.2 单调收敛定理 设 f 是从 X 映射到 $[-\infty,\infty]$ 的可测函数，满足 $f_n\uparrow f$ 和 $\int f_1\mathrm{d}\mu>-\infty$，那么 $\int f_n\mathrm{d}\mu\uparrow\int f\mathrm{d}\mu$.

证明 首先，验证 f 是可测的．对任意的 $c\in\mathbb{R}$，$f^{-1}((c,\infty])=\bigcup_{n\geqslant1}f_n^{-1}((c,\infty])\in\mathcal{S}$. 容易看出，所有开半直线 $(c,\infty]$ 构成的集生成了 $[-\infty,\infty]$ 的博雷尔 σ-代数(参见定理 3.2.6 之前的讨论)，那么根据定理 4.1.6，f 是可测的．

第二步，假设 $f_1\geqslant0$，那么根据命题 4.1.5，对每个 n，当 $m\to\infty$ 时，取简单函数 $f_{nm}\uparrow f_n$，令 $g_n:=\max(f_{1n},f_{2n},\cdots,f_{nn})$，那么每个 g_n 是简单函数且 $0\leqslant g_n\uparrow f$. 所以根据命题 4.1.5，$\int g_n\mathrm{d}\mu\uparrow\int f\mathrm{d}\mu$. 由于 $g_n\leqslant f_n\uparrow f$，因此，$\int f_n\mathrm{d}\mu\uparrow\int f\mathrm{d}\mu$.

第三步，假设 $f\leqslant0$，令 $g_n:=-f_n\downarrow -f:=g$，那么对所有的 n，$0\leqslant\int g\mathrm{d}\mu\leqslant\int g_n\mathrm{d}\mu<+\infty$ (中间的

不等式由引理 4.1.9 得到). 现在 $0 \leq g_1 - g_n \uparrow g_1 - g$, 所以 $\int g_1 - g_n \mathrm{d}\mu \uparrow \int g_1 - g\mathrm{d}\mu < +\infty$. 这些积分都是有限的, 根据定理 4.1.10, 我们可以从 $\int g_1 \mathrm{d}\mu$ 中抽取它们, 并且得到 $\int g_n \mathrm{d}\mu \downarrow \int g \mathrm{d}\mu$, 所以正如所期望的, 有 $\int f_n \mathrm{d}\mu \uparrow \int f \mathrm{d}\mu$.

在一般情况下, 有 $f_n^+ \uparrow f^+$ 和 $f_n^- \downarrow f^-$, 满足 $\int f^- \mathrm{d}\mu < +\infty$, 所以根据前面的情况, 有 $\int f_n^+ \mathrm{d}\mu \uparrow \int f^+ \mathrm{d}\mu$ 和 $+\infty \geq \int f_n^- \mathrm{d}\mu \downarrow \int f^- \mathrm{d}\mu \geq 0$, 因此 $\int f_n \mathrm{d}\mu \uparrow \int f \mathrm{d}\mu$. □

例如, 令 $f_n := -1_{[n,\infty)}$, 那么 $f_n \uparrow 0$, 但是 $\int f_n \mathrm{d}\mu = -\infty$, 不收敛到 0. 这说明了为什么假设 $\int f_1 \mathrm{d}\mu > -\infty$, 在单调收敛定理中是必需的. 存在一个满足 $f_n \downarrow f$, $\int f_1 \mathrm{d}\mu < +\infty$ 的单调收敛定理的对称形式.

对任意的实数 a_n, $\liminf_{n\to\infty} a_n := \sup_m \inf_{n \geq m} a_n$ 有定义(可能有限). 所以下面的引理, 尽管是单边不等式, 但也是相当一般的.

4.3.3 法图引理　设 f_n 是 X 上的任意非负可测函数, 那么 $\int \liminf f_n \mathrm{d}\mu \leq \liminf \int f_n \mathrm{d}\mu$.

131

证明　令 $g_n(x) := \inf\{f_m(x):\ m \geq n\}$, 那么 $g_n \uparrow \liminf f_n$. 因此, 根据单调收敛定理, $\int g_n \mathrm{d}\mu \uparrow \int \liminf f_n \mathrm{d}\mu$. 对所有的 $m \geq n$, $g_n \leq f_m$, 所以 $\int g_n \mathrm{d}\mu \leq \int f_m \mathrm{d}\mu$. 因此 $\int g_n \mathrm{d}\mu \leq \inf\left\{\int f_m \mathrm{d}\mu : m \geq n\right\}$. 当 $n \to \infty$ 时, 两边取极限完成证明. □

4.3.4 推论　假设 f_n 是非负可测函数, 满足对所有的 x, $f_n(x) \to f(x)$, 那么 $\int f \mathrm{d}\mu \leq \sup_n \int f_n \mathrm{d}\mu$.

例: 令 $\mu = \lambda$, $f_n = 1_{[n,n+1)}$, 那么对所有的 n, 当 $\int f_n \mathrm{d}\mu = 1$ 时, 对所有的 x, $f_n(x) \to f(x) := 0$. 这说明法图(Fatou)引理中的不等式可以是严格的. 这个例子也可帮助理解哪种不等式适用.

4.3.5 控制收敛定理　设 f_n, $g \in \mathcal{L}^1(X, \mathcal{S}, \mu)$, $|f_n(x)| \leq g(x)$ 且对所有的 x, $f_n(x) \to f(x)$, 那么 $f \in \mathcal{L}^1$ 且 $\int f_n \mathrm{d}\mu \to \int f \mathrm{d}\mu$.

证明　令 $h_n(x) := \inf\{f_m(x):\ m \geq n\}$, $j_n(x) := \sup\{f_m(x):\ m \geq n\}$, 那么 $h_n \leq f_n \leq j_n$. 因为 $h_n \uparrow f$ 和 $\int h_1 \mathrm{d}\mu \geq -\int |g| \mathrm{d}\mu > -\infty$, 所以有 $\int h_n \mathrm{d}\mu \uparrow \int f \mathrm{d}\mu$ (根据定理 4.3.2). 类似地考虑 $-j_n$, 有 $\int j_n \mathrm{d}\mu \downarrow \int f \mathrm{d}\mu$. 因为 $\int h_n \mathrm{d}\mu \leq \int f_n \mathrm{d}\mu \leq \int j_n \mathrm{d}\mu$, 所以得到 $\int f_n \mathrm{d}\mu \to \int f \mathrm{d}\mu$. □

例: $1_{[n,2n]} \to 0$, 但 $\int 1_{[n,2n]} \mathrm{d}\lambda = n \not\to 0$. 这说明了"控制"假设 $|f_n| \leq g \in \mathcal{L}^1$ 是有用的.

注: 求和可以看作是对计数测度(它给每个单元素集的测度为 1)的积分, 所以上面的收敛定理均可用于求和.

习题

1. 设$f_n \in \mathcal{L}^1(X, \mathcal{S}, \mu)$，满足对所有的$x$，当$n \to \infty$时，$f_n \geq 0$，$f_n(x) \to f_0(x)$且$\int f_n d\mu \to \int f_0 d\mu < \infty$，证明：$\int |f_n \to f_0| d\mu \to 0$. [提示：$(f_n - f_0)^- \leq f_0$；利用控制收敛定理.]

132

2. 在法图引理的叙述中，考虑用"lim sup"代替"lim inf"，用"$f_n \leq 0$"代替"$f_n \geq 0$"，用"$\geq$"代替"$\leq$"，证明：如果所有三个都改变了，结论仍然成立，但是举例说明如果一个或两个改变了，结论不成立.
3. 假设对所有的x，$f_n(x) \geq 0$，当$n \to \infty$时，$f_n(x) \to f_n(x)$. 假设$\int f_n d\mu$收敛到某个$c > 0$，证明：$\int f d\mu$的定义，但在区间$[0, c]$中不一定等于c. 举例说明$[0, c]$中的任意值都是可能的.
4. (a) 设$f_n := 1_{[0,n]}/n^2$，问是否存在一个控制这些f_n的可积函数g(对于$\mathbb{R}$上的勒贝格测度)？
   (b) 对$f_n := 1_{[0,n]}/(n\log n)$，$n \geq 2$，问题同(a).
5. 证明：当$n \to \infty$时，$\int_0^\infty \sin(e^x)/(1 + nx^2)dx \to 0$.
6. 证明：当$n \to \infty$时，$\int_0^1 (n\cos x)/(1 + n^2x^{3/2})dx \to 0$.
7. 在一个测度空间$(X, \mathcal{S}, \mu)$上，设f是一个实值可测函数且满足$\int f^2 d\mu < \infty$，g_n是一个可测函数满足对所有的x，$|g_n(x)| \leq f(x)$且对所有的x，$g_n(x) \to g(x)$，证明：当$n \to \infty$时，$\int (g_n + g)^2 d\mu \to 4\int g^2 d\mu < \infty$.
8. 通过考虑序列$g + f_n$和$g - f_n$，利用法图引理证明控制收敛定理.
9. 对$x > 1$，令$g(x) := 1/(x\log x)$，对某个常数$c_n \geq 0$和$[2, \infty)$的可测子集$A(n)$，令$f_n := c_n 1_{A(n)}$，证明或反证：如果对所有的x，有$f_n(x) \to 0$和$|f_n(x)| \leq g(x)$，那么当$n \to \infty$时，$\int_2^\infty f_n(x)dx \to 0$.
10. 设$f(x, y)$是二元实可测函数，有偏导数$\partial f/\partial x$，对$a < x < b$和$c \leq y \leq d$，此偏导数有界，其中c和d是有限数，是使得对某个$x \in (a, b)$，有$\int_c^d |f(x, y)| dy < \infty$，证明：对所有的$x \in (a, b)$，积分有限，且我们可以"在积分符号下求导"即，对$a < x < b$，$(d/dx)\int_c^d f(x,y)dy = \int_c^d \partial f(x, y)/\partial x dy$.
11. 如果$c = 0$和$d = +\infty$，且对某个$x = x_0$，$\int_0^\infty |\partial f(x,y)/\partial x| dy < \infty$，证明：对这个$x$，习题10的结论不一定成立. [提示：令$a = -1$，$b = 1$，$x_0 = 0$以及对$x \leq 1/(y+1)$，$f(x, y) = 0$.]
12. 设f_n和g_n是对测度μ的可积函数，满足$|f_n| \leq g_n$. 假设当$n \to \infty$时，$f_n(x) \to f(x)$，以及对几乎所有的x，$g_n(x) \to g(x)$，证明：如果$\int g_n d\mu \to \int g d\mu < \infty$，那么$\int f_n d\mu \to \int f d\mu$. [提示：参考习题8.]
13. 设$(X, \mathcal{S}, \mu)$是有限测度空间，即$\mu(X) < \infty$. X上的实值可测函数序列$\{f_n\}$称为依测度收敛(converge in measure)到f，如果对每个$\varepsilon > 0$，$\lim_{n\to\infty}\mu\{x: |f_n(x) - f(x)| > \varepsilon\} = 0$. 证明：如果$f_n \to f$依测度且对某个可积函数$g$及所有的$n$，满足$|f_n| \leq g$，则$\int |f_n - f| d\mu \to 0$.

133

14. (a) 如果$f_n \to f$依测度且$\int f_n d\mu \to \int f d\mu < \infty$，证明：$\int |f_n - f| d\mu \to 0$. [提示：参考习题1.]
    (b) 如果$f_n \to f$依测度且$\int |f_n| d\mu \to \int |f| d\mu < \infty$，证明：$\int |f_n - f| d\mu \to 0$.
15. 对一个有限测度空间$(X, \mathcal{S}, \mu)$，可积函数集合$\mathcal{F}$称为一致可积的(uniformly integrable)，当且仅当$\sup\left\{\int |f| d\mu : f \in \mathcal{F}\right\} < \infty$，且对任意的$\varepsilon > 0$，存在$\delta > 0$，使得对$\mu(A) < \delta$，及对任意的$f \in \mathcal{F}$，有$\int_A |f| d\mu < \varepsilon$. 证明：可积函数序列$\{f_n\}$满足$\int |f_n - f| d\mu \to 0$当且仅当$f_n \to f$依测度且$f_n$是一致可积的.

4.4 乘积测度

对任意的 $a\leqslant b$ 和 $c\leqslant d$，$\mathbb{R}^2$ 中的矩形 $[a, b]\times[c, d]$ 有面积 $(d-c)(b-a)$，即是边界长度的乘积．对更一般的集合，面积的这种定义是很熟悉的．本节将把面积更一般化地定义为测度，并将定义任意两个 σ-有限测度的笛卡儿积，这两个 σ-有限测度代替了两个实数轴上的长度．从而这个乘积将能扩张到两个以上因子，例如，$\mathbb{R}^3$ 中的体积作为测度．

设 $(X, \mathcal{B}, \mu)$ 和 $(Y, \mathcal{C}, \nu)$ 为任意两个测度空间，在 $X\times Y$ 中设 $\mathcal{R}$ 是满足 $B\in\mathcal{B}$ 和 $C\in\mathcal{C}$ 的所有"矩形" $B\times C$ 构成的集类．对这种集合，令 $\rho(B\times C):=\mu(B)\nu(C)$，其中（在这种情况下），规定 $0\cdot\infty:=\infty\cdot 0:=0$．根据命题 3.2.2，$\mathcal{R}$ 是一个半环．

4.4.1 定理　*在 $\mathcal{R}$ 上 ρ 是可数可加的．*

证明　假设在 $\mathcal{R}$ 中，$B\times C=\bigcup_n B(n)\times C(n)$，其中集合 $B(n)\times C(n)$ 是互不相交的，对所有的 n，$B(n)\in\mathcal{B}$，$C(n)\in\mathcal{C}$，所以对每个 $x\in X$，$y\in Y$，$1_B(x)1_C(y)=\sum_n 1_{B(n)}(x)1_{C(n)}(y)$．那么根据可数可加性，积分 $\mathrm{d}\nu(y)$ 给出了对每个 x，$1_B(x)\nu(C)=\sum_n 1_{B(n)}(x)\nu(C(n))$．现在根据可加性（命题 4.1.8）和单调收敛性（定理 4.3.2），对 $\mathrm{d}\mu(x)$ 求积分，有 $\mu(B)\nu(C)=\sum_n \mu(B(n))\nu(C(n))$．　□

设 $\mathcal{A}$ 是由 $\mathcal{R}$ 生成的环，那么 $\mathcal{A}$ 由 $\mathcal{R}$ 的有限多个不相交的元素的所有并组成（命题 3.2.3）．因为 $X\times Y\in\mathcal{R}$，$\mathcal{A}$ 是一个代数，对任意不相交的 $C_j\in\mathcal{R}$ 和有限的 n，令 $\rho\left(\bigcup_{1\leqslant j\leqslant n}C_j\right)>:=\sum_{1\leqslant j\leqslant n}\rho(C_j)$，根据命题 3.2.4 和定理 4.4.1，这里 ρ 是明确定义的而且在 $\mathcal{A}$ 上是可数可加的．从而 ρ 可以扩张为乘积 σ-代数 $\mathcal{B}\otimes\mathcal{C}$ 上的一个可数可加测度（命题 4.1.7 之前的定义），$\mathcal{B}\otimes\mathcal{C}$ 由 $\mathcal{R}$ 或 $\mathcal{A}$ 生成，但是在一般情况下，这种扩张不唯一．下一步将给出扩张唯一且能写成累次积分的条件．为此，下面的概念很有用．一个集类 $\mathcal{M}$ 称为单调类（monotone class）当且仅当只要 $M_n\in\mathcal{M}$ 且 $M_n\downarrow M$ 或 $M_n\uparrow M$，就有 $M\in\mathcal{M}$．例如，任意的 σ-代数是一个单调类，但一般地，一个拓扑不是一个单调类（开集的一个无限交通常不是开的）．对任意的集 X，2^X 显然是一个单调类．任意单调类集合的交是一个单调类．因此，对任意的集类 $\mathcal{D}$，存在一个包含 $\mathcal{D}$ 的最小单调类． 134

4.4.2 定理　*如果 $\mathcal{A}$ 是集合 X 的子集构成的代数，那么包含 $\mathcal{A}$ 的最小单调类 $\mathcal{M}$ 是一个 σ-代数．*

证明： 令 $\mathcal{N}:=\{E\in\mathcal{M}: X\setminus E\in\mathcal{M}\}$，那么 $\mathcal{A}\subset\mathcal{N}$ 且 $\mathcal{N}$ 是一个单调类，所以 $\mathcal{N}=\mathcal{M}$．对每个 $A\subset X$，令 $\mathcal{M}_A:=\{E: E\cap A\in\mathcal{M}\}$，那么对 $\mathcal{A}$ 中的每个 A，$\mathcal{A}\subset\mathcal{M}_A$ 且 $\mathcal{M}_A$ 是一个单调类，所以 $\mathcal{M}\subset\mathcal{M}_A$．从而对每个 $E\in\mathcal{M}$，$\mathcal{M}_E$ 是包含 $\mathcal{A}$ 的一个单调类，所以 $\mathcal{M}\subset\mathcal{M}_E$．因此 $\mathcal{M}$ 是一个代数．作为一个单调类，还是一个 σ-代数．　□

下面的定理说明可测集的示性函数的积分顺序可以交换（在乘积 σ-代数中）．对于定理 4.4.4 中乘积测度的构造和更一般函数的积分的交换，这将是最重要的一步．

4.4.3 定理　*假设 $\mu(X)<+\infty$ 和 $\nu(Y)<+\infty$，令*

$$\mathcal{F}:=\left\{E\subset X\times Y:\int\left[\int 1_{E(x,y)}\mathrm{d}\mu(x)\right]\mathrm{d}\nu(y)=\int\left[\int 1_{E(x,y)}\mathrm{d}\nu(y)\right]\mathrm{d}\mu(x)\right\}.$$

那么 $\mathcal{B}\otimes\mathcal{C}\subset\mathcal{F}$.

证明 $\mathcal{F}$的定义暗示了$\mathcal{F}$中的所有积分有定义，因此每个被积函数是可测的．注意到，这种可测性在下面的每一步证明中都是成立的．如果对某个 $B \in \mathcal{B}$和 $C \in \mathcal{C}$，$E = B \times C$，那么

135

$$\iint 1_E \mathrm{d}\mu \mathrm{d}\nu = \mu(B)\int 1_C \mathrm{d}\nu = \mu(B)\nu(C) = \iint 1_E \mathrm{d}\nu \mathrm{d}\mu.$$

因此$\mathcal{R} \subset \mathcal{F}$. 如果 $E_n \in \mathcal{F}$，且 $E_n \downarrow E$ 或 $E_n \uparrow E$，那么根据单调收敛性(定理 4.3.2)，利用有限性，有 $E \in \mathcal{F}$. 因此$\mathcal{F}$是一个单调类．同理，$\mathcal{F}$中集合的任意有限不交集的并还在$\mathcal{F}$中．从而$\mathcal{A} \subset \mathcal{F}$，因此根据定理 4.4.2，$\mathcal{B} \otimes \mathcal{C} \subset \mathcal{F}$. □

4.4.4 乘积测度存在定理 设$(X, \mathcal{B}, \mu)$和$(Y, \mathcal{C}, \nu)$是两个 σ-有限测度空间，那么 ρ 在$\mathcal{B} \otimes \mathcal{C}$上的扩张测度是唯一的，且满足对所有的 $E \in \mathcal{B} \otimes \mathcal{C}$，

$$\rho(E) = \iint 1_E(x,y)\mathrm{d}\mu(x)\mathrm{d}\nu(y) = \iint 1_E(x,y)\mathrm{d}\nu(y)\mathrm{d}\mu(x).$$

证明 首先，假设 μ 和 ν 是有限的，令 $\alpha(E) := \iint 1_E(x, y)\mathrm{d}\mu(x)\mathrm{d}\nu(y)$，$E \in \mathcal{B} \otimes \mathcal{C}$. 则根据定理 4.4.3，$\alpha$ 有定义且积分顺序可以交换．根据命题 4.1.8，现在 α 是有限可加的(对$\mathcal{B} \otimes \mathcal{C}$中任意有限多个不交集)．其次，所有被积函数是可测的．从而根据单调收敛性(定理 4.3.2)，α 是可数可加的．对 ρ 在$\mathcal{B} \otimes \mathcal{C}$上的任意其他扩张测度 β，满足 $\alpha = \beta$ 的集类是包含$\mathcal{A}$的一个单调类，因此也包含$\mathcal{B} \otimes \mathcal{C}$. 所以定理对有限测度成立．

一般地，令 $X = \bigcup_m B_m$，$Y = \bigcup_n C_n$，其中 B_m 和 C_n 分别在 X 和 Y 中互不相交，满足对所有的 m 和 n，有 $\mu(B_m) < \infty$ 和 $\nu(C_n) < \infty$. 令 $E \in \mathcal{B} \otimes \mathcal{C}$且 $E(m, n) := E \cap (B_m \times C_n)$，那么对每个 m 和 n，根据有限的情况，

$$\iint 1_{E(m,n)} \mathrm{d}\mu \mathrm{d}\nu = \iint 1_{E(m,n)} \mathrm{d}\nu \mathrm{d}\mu.$$

这个等式可以对所有的 m 和 n 求和(根据引理 3.1.2，以任意的顺序)．根据可数可加性和单调收敛性，有

$$\alpha(E) := \iint 1_E \mathrm{d}\mu \mathrm{d}\nu = \iint 1_E \mathrm{d}\nu \mathrm{d}\mu, \quad \text{对任意的 } E \in \mathcal{B} \otimes \mathcal{C}.$$

则根据单调收敛性，α 是有限可加、可数可加的，并且是一个在$\mathcal{A}$上等于 ρ 的测度．如果 β 是 ρ 在$\mathcal{B} \otimes \mathcal{C}$上的任意其他扩张测度，那么对任意的 $E \in \mathcal{B} \otimes \mathcal{C}$，

$$\beta(E) = \sum_{m,n} \beta(E(m,n)) = \sum_{m,n} \alpha(E(m,n)) = \alpha(E),$$

136

所以扩张是唯一的． □

例： 令 c 是计数测度，λ 是 $I := [0, 1]$上的勒贝格测度．在 $I \times I$ 中，令 D 是对角集合$\{(x, x): x \in I\}$，那么 D 是可测的(它是闭的且 I 是第二可数的，所以命题 4.1.7 适用)，但是$\iint 1_D \mathrm{d}\lambda \mathrm{d}c = 0 \neq 1 = \iint 1_D \mathrm{d}c \mathrm{d}\lambda$. 这表明 σ-有限性在定理 4.4.4 中是有用的(c 不是 σ-有限的)．

$\mathcal{B} \otimes \mathcal{C}$上的测度 ρ 称为是一个乘积测度 $\mu \times \nu$. 下面给出一个关于乘积测度积分的重要定理.

4.4.5 定理(Tonelli-Fubini) 设$(X, \mathcal{B}, \mu)$和$(Y, \mathcal{C}, \nu)$是两个 σ-有限空间，f 是一个从 $X \times Y$ 映射到$[0, \infty]$的关于$\mathcal{B} \otimes \mathcal{C}$可测的函数，或 $f \in \mathcal{L}^1(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \times \nu)$，那么

$$\int f \mathrm{d}(\mu \times \nu) = \iint f(x,y)\mathrm{d}\mu(x)\mathrm{d}\nu(y) = \iint f(x,y)\mathrm{d}\nu(y)\mathrm{d}\mu(x).$$

这里$\int f(x,y)\,\mathrm{d}\mu(x)$对$\nu$-几乎所有的$y$有定义，$\int f(x,y)\,\mathrm{d}\nu(y)$对$\mu$-几乎所有的$x$有定义．

证明：由于只有函数几乎处处有定义时，其积分有定义(如在命题4.3.1和其之后)．对非负简单函数f，由定理4.4.4和积分的可加性(命题4.1.8)可知，定理成立．对非负可测函数f，由单调收敛性(命题4.1.5和定理4.3.2)可知，定理成立．对$f\in\mathcal{L}^1(X\times Y,\ \mathcal{B}\otimes\mathcal{C},\ \mu\times\nu)$，且对$f^+$和$f^-$定理成立．从而，对几乎所有的$y$，$\int f^+(x,y)\,\mathrm{d}\mu(x)<\infty$(关于$\nu$)，同理对$f^-$以及$\mu$和$\nu$互相交换也成立．对$\nu$-几乎所有的$y$，$\int|f(x,y)|\,\mathrm{d}\mu(x)<\infty$，则根据定理4.1.10，

$$\int f(x,y)\,\mathrm{d}\mu(x)=\int f^+(x,y)\,\mathrm{d}\mu(x)-\int f^-(x,y)\,\mathrm{d}\mu(x),$$

且这三个积分都是有限的．下一步，正如命题4.3.1中那样，在ν-测度为0的集合上关于ν求积分，其中在此集合上f^+和f^-的积分无限是无关紧要的．因此根据定理4.1.10，$\iint f(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)$有定义、有限且等于

$$\iint f^+(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)-\iint f^-(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y).$$

从而定理对f^+和f^-成立意味着对f成立．□　137

注：要证明$f\in\mathcal{L}^1(X\times Y,\ \mathcal{B}\otimes\mathcal{C},\ \mu\times\nu)$，只需要证明$f$是$\mathcal{B}\otimes\mathcal{C}$可测的，且$\iint|f|\,\mathrm{d}\mu\mathrm{d}\nu<+\infty$或$\iint|f|\,\mathrm{d}\nu\mathrm{d}\mu<+\infty$．

例(a)：设$X=Y=\mathbb{N}$，$\mu=\nu=$计数测度，因此，对$f:\ \mathbb{N}\mapsto\mathbb{R}$，有$\int f\mathrm{d}\mu=\int f(n)\,\mathrm{d}\mu(n)=\sum_n f(n)$，其中$f\in\mathcal{L}^1(\mu)$当且仅当$\sum_n|f(n)|<+\infty$．(对计数测度，$\sigma$-代数是$2^{\mathbb{N}}$，从而所有的函数是可测的．)在$\mathbb{N}\times\mathbb{N}$上，对所有的$n\in\mathbb{N}$，令$g(n,\ n):=1$，$g(n+1,\ n):=-1$，对$m\neq n$和$m\neq n+1$，令$g(m,\ n):=0$(参见图4-4)．那么$g$是有界的且在$X\times Y$上可测，$\iint g(m,n)\,\mathrm{d}\mu(m)\,\mathrm{d}\nu(n)=(1-1)+(1-1)+\cdots=0$，但是$\iint g(m,n)\,\mathrm{d}\nu(n)\,\mathrm{d}\mu(m)=1+(1-1)+(1-1)+\cdots=1$．因此，这个积分不能交换次序($g^+$和$g^-$都有无限积分)．

0	0	0	0	0	1	−1	0	0
0	0	0	0	1	−1	0	0	0
0	0	0	1	−1	0	0	0	0
0	0	1	−1	0	0	0	0	0
0	1	−1	0	0	0	0	0	0
和	1	0	0	0	0	0	0	0

图　4-4

例(b)：对$x\in\mathbb{R}$和$t>0$，令

$$f(x,t):=(2\pi t)^{-1/2}\exp(-x^2/(2t)),$$

$g(x,\ t):=\partial f/\partial t$，那么$\partial f/\partial x=-xf/t$和$\partial^2/\partial x^2=(x^2t^{-2}-t^{-1})f=2g$．因此，$f$满足偏微分方程$2\partial f/\partial t=\partial^2f/\partial x^2$，称为热传导方程(heat equation)．(事实上，f称为热传导方程的一个“基本解”：

Schwartz, 1966, p. 145.)对每个 $t>0$，有 $\int_{-\infty}^{\infty} f(x,t)\,\mathrm{d}x=1$(其中 $\mathrm{d}x:=\mathrm{d}\lambda(x)$)，根据(习题6中讨论)极坐标有

$$\left(\int_0^{\infty}\exp(-x^2/2)\,\mathrm{d}x\right)^2=(\pi/2)\int_0^{\infty} r\cdot\exp(-r^2/2)\,\mathrm{d}r=\pi/2.$$

(这里$f(\cdot,t)$称为"正态"或"高斯"概率密度．这一函数在第9章中将起着重要作用．)由于当 $|x|\to\infty$ 时，$\partial f/\partial x\to 0$，所以对任意的 $s>0$，

$$\int_{-\infty}^{\infty}\int_s^{\infty} g(x,t)\,\mathrm{d}t\mathrm{d}x=\int_{-\infty}^{\infty}-f(x,s)\,\mathrm{d}x=-1,$$

但

138

$$\int_s^{\infty}\int_{-\infty}^{\infty} g(x,t)\,\mathrm{d}x\mathrm{d}t=\int_s^{\infty}\partial f/\partial x\Big|_0^{\infty}\mathrm{d}t=0$$

(根据洛必达法则)．因此，这里的积分次序不能交换，g^+和g^-都是不可积的．关于这个矛盾的一些观点来自于施瓦茨(Schwartz)的分布理论，在这一理论中尽管函数 $2\partial f/\partial t$ 和$\partial^2 f/\partial x^2$ 是光滑的且对 $t>0$ 是相等的，但它们的差不是0，而是集中在(0，0)处的测度．参考 Hörmander(1983，p. 80—81)．

设对 $j=1,\cdots,n$，$\mathcal{S}_j$是X_j子集的σ-代数，则使得每个坐标函数 x_j 可测的最小σ-代数定义为乘积σ-代数$\mathcal{S}_1\otimes\cdots\otimes\mathcal{S}_n$．很容易看出，这与前面对 $n=2$ 的定义及对 $n\geqslant 2$ 的定义一致，通过对 n 作归纳，$\mathcal{S}_1\otimes\cdots\otimes\mathcal{S}_n$是 X 子集的最小σ-代数，X 包含满足 $A_j\in\mathcal{S}_j(j=1,\cdots,n)$的所有集合 $A_1\times\cdots\times A_n$．正如对积拓扑，连续函数称为联合连续，对积$\sigma$-代数，可测函数将称为联合可测(jointly measurable)．

4.4.6 定理 *设$(X_j,\mathcal{S}_j,\mu_j)$是$\sigma$-有限测度空间，$j=1,\cdots,n$，那么在 $X=X_1\times\cdots\times X_n$ 上关于乘积σ-代数$\mathcal{S}$存在唯一的测度μ，满足对任意的 $A_j\in\mathcal{S}_j$，$j=1,\cdots,n$，$\mu(A_1\times\cdots\times A_n)=\mu_1(A_1)\mu_2(A_2)\cdots\mu_n(A_n)$，或者如果任意的$\mu_j(A_j)=0$，即使另一个为$+\infty$，$\mu(A_1\times\cdots\times A_n)=0$．如果$f$是 X 上的非负联合可测函数，或如果$f\in\mathcal{L}^1(X,\mathcal{S},\mu)$，那么*

$$\int f\mathrm{d}\mu=\int\cdots\int f(x_1,\cdots,x_n)\,\mathrm{d}\mu_1(x_1)\cdots\mathrm{d}\mu_n(x_n),$$

其中对$f\in\mathcal{L}^1(X,\mathcal{S},\mu)$，累次积分是在对$\mu_n$-几乎所有的 x_n 的意义下递归地"从外部"定义的，关于其他变量的累次积分是有定义且有限的，所以除了在一个μ_{n-1}测度为0的集合上(可能依赖于x_n)，对前 $n-2$ 个变量的累次积分是有定义且有限的，等等．如果这个积分是任意次序的，那么有相同的结论成立．

证明 根据定理4.4.4和定理4.4.5，以及对 n 作归纳可得证明．□

关于定理4.4.6的最著名的一个例子是$\mathbb{R}^n$上的勒贝格测度λ^n，它是满足在$\mathbb{R}$上对每个j，μ_j为勒贝格测度λ的一个乘积，那么λ是长度，λ^2是面积，λ^3是体积，等等．

139

习题

1. 设$(X,\mathcal{B},\mu)$和$(Y,\mathcal{C},\nu)$是两个σ-有限测试空间，$f\in\mathcal{L}^1(X,\mathcal{B},\mu)$，$g\in\mathcal{L}^1(Y,\mathcal{C},\nu)$，$h(x,y):=f(x)g(y)$，证明：$h\in\mathcal{L}^1(X\times Y,\mathcal{B}\otimes\mathcal{C},\mu\times\nu)$，且 $\int h\mathrm{d}(\mu\times\nu)=\int f\mathrm{d}\mu\int g\mathrm{d}\nu$.

2. 设$(X,\mathcal{B},\mu)$是σ-有限测度空间，f是 X 上的一个非负可测函数，证明：对 $\lambda:=$勒贝格测度，$\int f\mathrm{d}\mu=(\mu\times\lambda)\{(x,y):0<y<f(x)\}$("积分是曲线下的面积")．

3. 设$(X, \mathcal{B}, \mu)$是σ-有限的，f是X上的任意可测实值函数，证明：$(\mu\times\lambda)\{(x, y): y=f(x)\}=0$(一个实可测函数的图像的测度为0).
4. 设$(X, \leqslant)$是一个不可数的良序集，满足对任意的$y\in X$，$\{x\in X: x<y\}$是可数的. 对任意的$A\subset X$，如果A是可数的，令$\mu(A)=0$，如果$X\setminus A$是可数的，令$\mu(A)=1$证明：μ是一个定义在σ-代数上的测度. 通过$T:=\{(x, y)\in X\times X: y<x\}$定义$T$，"序数三角形". 求累次积分$\iint 1_T(x,y)\mathrm{d}\mu(y)\mathrm{d}\mu(x)$和$\iint 1_T(x, y)\mathrm{d}\mu(x)\mathrm{d}\mu(y)$的值. 怎样使得结果与乘积测度定理(4.4.4和4.4.5)一致的?
5. 对$\mathbb{R}^2$上的乘积测度$\lambda\times\lambda$(平面上通常的勒贝格测度)，很容易证明平行于坐标轴的矩形有等于它们通常的面积(面积为边界的乘积)的测度$\lambda\times\lambda$. 证明此结论对矩形而言不一定是坐标轴的平行面.
6. *极坐标*(polar coordinates). 设T是从$X:=[0, \infty)\times[0, 2\pi)$映上到$\mathbb{R}^2$由$T(r, \theta):=(r\cdot\cos\theta, r\cdot\sin\theta)$定义的函数，证明：$T$在$[0, \infty)\times[0, 2\pi)$上是1-1的. 令$\sigma$为$(0, \infty)$上由$\sigma(A):=\int_A r\mathrm{d}\lambda(r):=\int 1_A(r)r\mathrm{d}\lambda(r)$定义的测度，在$X$上$\mu:=\sigma\times\lambda$，证明：像测度$\mu\circ T^{-1}$是$\mathbb{R}^2$上的勒贝格测度$\lambda^2:=\lambda\times\lambda$. [提示：当$T^{-1}(B)$是矩形时，证明$\lambda^2(B)=\mu(T^{-1}(B))$，其中$s<r\leqslant t$和$\alpha<\theta\leqslant\beta$($B$是环面的一个扇区). 可以通过微积分或证明当$(t-s)/s$和$\beta-\alpha$很小时，$B$可以由习题5中的矩形从内部和外部来逼近；聚集集合$B$使之变大. 证明这种集合$B$的不交集的有限并是一个环，这个环生成$\mathbb{R}^2$中可测集的$\sigma$-代数(利用命题4.1.7). (不允许使用雅可比定理.)]
7. 对$(0, \infty)$上的一个实值可测函数f，令$P(f):=\{p\in(0, \infty): |f|^p\in\mathcal{L}^1(\mathbb{R}, \mathcal{B}, \lambda)\}$，其中$\mathcal{B}$是一个博雷尔$\sigma$-代数，证明：对$(0, \infty)$的每个子区间$J$，$J$的左边和右边可能是开的或闭的，存在某个$f$，使得$P(f)=J$. [提示：在$(0, 1]$和$[1, \infty)$上考虑函数，比如$x^a|\log x|^b$，其中$a=-1$时，$b=0$.]

140

8. 设$B_k(0, r):=\{x\in\mathbb{R}^k: x_1^2+\cdots+x_k^2\leqslant r^2\}$为$\mathbb{R}^k$中的一个半径为$r$的球，$v_k:=\lambda^k(B_k(0, 1))$($k$-维单位球的体积).
   (a)证明：对任意的$r\geqslant 0$，$\lambda^k(B_k(0, r))=v_k r^k$.
   (b)对所有的k，求ν_k的值. [提示：从已知的ν_1和ν_2的值人手，由k到$k+2$对$k=1, 2, \cdots$作归纳，用极坐标代替$\mathbb{R}^{k+2}=\mathbb{R}^k\times\mathbb{R}^2$中的$x_{k+1}$和$x_{k+2}$].
9. 令S^{k-1}为$\mathbb{R}^k$中的单位球($B_k(0, 1)$的边界)，延续习题8，$x\mapsto(|x|, x/|x|)$给出了一个从$\mathbb{R}^k\setminus\{0\}$到$(0, \infty)\times S^{k-1}$的一一对应的映射$T$，证明：像测度$\lambda^k\circ T^{-1}$是一个乘积测度，在$(0, \infty)$上，此测度可以由$\rho_k(A):=\int_A r^{k-1}\mathrm{d}r$，任意的博雷尔集$A\subset(0, \infty)$和$S^{k-1}$上的某个测度$\omega_k$给出. 并找出全部质量$\omega_k(S^{k-1})$. [提示：对任意的博雷尔集$A\subset(0, \infty)$和$B\subset S^{k-1}$，令$\gamma_k(A, B):=\lambda^k(\{x: |x|\in A, x/|x|\in B\})$，通过$\omega_k(B):=k\gamma_k((0, 1), B)$定义$\omega_k$，由简单的集合$A$开始再到处理一般的博雷尔集合，证明$\gamma_k(A, B)=\rho_k(A)\omega_k(B)$.]
10. 证明：α(定义见定理4.4.4的证明)是一个可数可加的测度，即使μ和ν不是σ-有限的.
11. 如果对指标集I中的所有i，$(X_i, \mathcal{S}_i, \mu_i)$是测度空间，其中集合$X_i$是不相交的，这些测度空间的*直和*(direct sum)定义为$X=\bigcup_i X_i$，令$\mathcal{S}:=\{A\subset X$: 对所有的$i$, $A\cap X_i\in\mathcal{S}_i\}$，对每个$A\in\mathcal{S}$，令$\mu(A):=\sum_i\mu(A\cap X_i)$，证明：$(X, \mathcal{S}, \mu)$是一个测度空间.
12. 一个测度空间称为是*可局部化的*，当且仅当它可以表示为有限测度空间的直和(参考3.3节). 证明：
   (a)任意σ-有限测度空间是可局部化的.
   (b)任意σ-有限测度空间的直和是可局部化的.
13. 考虑具有博雷尔σ-代数的单位正方形I^2，对每个$x\in I:=[0, 1]$，令I_x为垂直区间$\{(x, y): 0\leqslant y\leqslant 1\}$，设$\mu$为$I^2$上的测度，$I^2$由每个$I_x$上的一维勒贝格测度的直和给出. 类似地，令$J_y:=\{(x, y): 0\leqslant x\leqslant 1\}$，

ν 为每个 J_y 上的一维勒贝格测度的直和，$\mathcal{B}$为 I^2 中关于 μ 和 ν 可测的集类，证明或反证：$(I^2, \mathcal{B}, \mu+\nu)$ 是可局部化的．建议：参考习题4. 采用连续统假设(附录 A.3).

[141] 14. (Bledsoe-Morse 乘积测度). 给定两个 σ-有限测度空间 $(X, \mathcal{S}, \mu)$ 和 $(Y, \mathcal{T}, \nu)$，令 $\mathcal{N}(\mu, \nu)$ 是所有 $A\subset X\times Y$ 组成的集类，使得对 ν-几乎所有的 y，有 $\int 1_A(x, y)\,d\mu(x) = 0$，且对 μ-几乎所有的 x，$\int 1_A(x,y)\,d\nu(y) = 0$(对 y 和 x 的其他值，积分可能无限). 证明：乘积 σ-代数$\mathcal{S}\otimes\mathcal{T}$上的乘积测度$\mu\times\nu$ 可以扩张为 σ-代数$\mathcal{U}=\mathcal{U}(\mu, \nu)$上的测度 ρ，σ-代数$\mathcal{U}=\mathcal{U}(\mu, \nu)$由$\mathcal{S}\otimes\mathcal{T}$和$\mathcal{N}(\mu, \nu)$生成，且在$\mathcal{N}(\mu, \nu)$上，$\rho=0$，从而以$\mathcal{U}(\mu, \nu)$代替$\mathcal{S}\otimes\mathcal{T}$后，定理 4.4.5 成立．[提示：$\mathcal{N}(\mu, \nu)$是一个遗传 σ-环，如 3.3 节的习题 9 中定义的．]

15. 对具有博雷尔 σ-代数和勒贝格测度 λ 的 $I:=[0, 1]$，取具有乘积测度(体积)$\lambda^3=\lambda\times\lambda\times\lambda$ 的立方体 I^3，当 $y\neq z$ 时，令 $f(x, y, z):=1/\sqrt{|y-z|}$；当 $y=z$ 时，令 $f(x, y, z):=+\infty$，证明：f 对 λ^3 是可积的，但是对每个$z\in I$，满足$\int f(x, y, z)\,d\lambda(x)=+\infty$ 的 y 组成的集合是非空的且依赖于 z.

*4.5 丹尼尔-斯通积分

既然积分关于测度是有限的，则其过程是可逆的，在这一意义下，给定具有适当性质的“积分”算子，可以证明此积分可以表达为关于某个测度的积分．

设$\mathcal{L}$是由集合 X 上的实值函数组成的非空集类，则$\mathcal{L}$是一个实向量空间(real vector space)当且仅当对所有的$f, g\in\mathcal{L}$和 $c\in\mathbb{R}$，有 $cf+g\in\mathcal{L}$. 令$f\vee g:=\max(f, g)$，$f\wedge g:=\min(f, g)$. 一个函数向量空间$\mathcal{L}$称为向量格(vector lattice)当且仅当对所有的$f, g\in\mathcal{L}$，有$f\vee g\in\mathcal{L}$且$f\wedge g\equiv-(-f\vee -g)\in\mathcal{L}$.

例：对任意的测度空间$(X, \mathcal{S}, \mu)$，所有 μ-可积实函数组成的集合$\mathcal{L}^1(X, \mathcal{S}, \mu)$是一个向量格，所有 μ-简单函数组成的集合也是一个向量格．对任意的拓扑空间$(X, \mathcal{T})$，X 上的所有有界连续实值函数的集类$C_b(X, \mathcal{T})$是一个向量格．换句话说，令 C^1 为所有导数处处存在且连续的函数 f: $\mathbb{R}\mapsto\mathbb{R}$ 组成的集合，则 C^1 是一个向量空间但不是一个格．

定义 给定一个集合 X 和 X 上的实函数的一个向量格，一个**准整数**(pre-integral)是由$\mathcal{L}$映射到 $\mathbb{R}$ 的函数 I，I 满足以下条件：

(a)I 是线性的，即对所有的 $c\in\mathbb{R}$ 和 $f, g\in\mathcal{L}$，$I(cf+g)=cI(f)+I(g)$.

(b)I 是非负的，也就是只要$f\in\mathcal{L}$和 $f\geqslant 0$(在 X 的任何点处)，就有 $I(f)\geqslant 0$.

[142] (c)对所有的 x，只要$f_n\in\mathcal{L}$和 $f_n(x)\downarrow 0$，就有 $I(f_n)\downarrow 0$.

注：对任意的函数向量空间$\mathcal{L}$，常函数 0 属于$\mathcal{L}$. 因此，如果$\mathcal{L}$是一个向量格，那么对任意的 $f\in\mathcal{L}$，函数$f^+:=\max(f, 0)$和$f^-:=-\min(f, 0)$均属于$\mathcal{L}$，非负函数也属于$\mathcal{L}$. 所以在$\mathcal{L}$中存在足够多的非负函数，使得(b)有意义.

例：令$\mathcal{L}=C[0, 1]$，即由$[0, 1]$上的所有连续实值函数构成的空间，则$\mathcal{L}$是一个向量格．令 $I(f)$为经典的黎曼积分，正如微积分中，$I(f):=\int_0^1 f(x)\,dx$，那么显然 I 是非负线性函数．如果$f_n\in C[0, 1]$且$f_n\downarrow 0$，那么根据迪尼定理 2.2.10，$f_n\downarrow 0$ 在$[0, 1]$上是一致成立的．因此，$I(f_n)\downarrow 0$ 且(c)成立，从而 I 是准整数．类似地，对任意的紧拓扑空间 K 和 K 上的连续函数构成的空间 $C(K)$，如果 I 是 $C(K)$上的非负线性函数，那么 I 是一个准整数(这将在 7.4 节中讨论).

在本节的其余部分中，假设给定集合 X，X 上实函数的一个向量格$\mathcal{L}$，以及$\mathcal{L}$上的一个准整数 I，对$\mathcal{L}$中满足$f\leqslant g$(即对所有的 x，$f(x)\leqslant g(x)$)的任意两个函数 f 和 g，令

$$[f,g) := \{\langle x,t\rangle \in X\times\mathbb{R} : f(x)\leqslant t < g(x)\}.$$

设$\mathcal{S}$是$\mathcal{L}$中所有对$f\leqslant g$，由$[f, g)$构成的集类．在$\mathcal{S}$上通过$\nu([f, g)) := I(g-f)$定义ν，因此，如果$g\geqslant 0$，那么$I(g)=\nu([0, g))$，从而对任意的$f\in\mathcal{L}$，$I(f)=\nu([0, f^+))-\nu([0, f^-))$.

下面的定理将为丹尼尔-斯通(Daniell-Stone)定理(4.5.2)提供一个行之有效的方法．

4.5.1 定理(A. C. Zaanen)　*ν扩张为由$\mathcal{S}$生成的σ-代数$\mathcal{T}$上的一个可数可加测度．*

证明　首先证明$\mathcal{S}$是一个半环，ν在其上是明确定义的且是可数可加的．利用命题3.2.4，将推得此定理成立．

正如在命题3.2.1中那样，$\mathcal{S}$是一个半环：对$f=g$，$\varnothing\in\mathcal{S}$，且就像对$\mathbb{R}$中的区间那样，对$\mathcal{S}$中的任意函数$f\leqslant g$和$h\leqslant j$，$[f, g)\cap[h, j)=[f\vee h, f\vee h\vee(g\wedge j))$，同理，

$$[f,g)\setminus[h,j) = [f,f\vee(g\wedge h))\cup[g\wedge(j\vee f),g),$$

是$\mathcal{S}$中两个不交集的并．

假设$[f, g)=[h, j)$，那么对任意x，如果区间$[f(x), g(x))$是非空的，它等价于$[h(x), j(x))$，则有$f(x)=h(x)$，$g(x)=j(x)$且$(g-f)(x)=(j-h)(x)$. 另一方面，如果$[f(x), g(x))$是空的，那么$[h(x), j(x))$是空的，且$f(x)=g(x)$，$h(x)=j(x)$，所以$(g-f)(x)=0=(j-h)(x)$，从而$g-f\equiv j-h\in\mathcal{L}$且$I(g-f)=I(j-h)$，因此，$\nu$在$\mathcal{S}$上是明确定义的．

143

对可数可加性而言，如果$[f, g)=\bigcup_n[f_n, g_n)$，其中$f$, f_n, g和g_n是$\mathcal{L}$中的函数且集合$[f_n, g_n)$是不相交的，则对每个x，有$[f(x), g(x))=\bigcup_n[f_n(x), g_n(x))$，其中区间$[f_n(x), g_n(x))$是不相交的．正如对区间$(a, b]$那样，根据对称性，对区间$[a, b)$，区间的长度是可数可加的(对$G(x)\equiv x$，定理3.1.3)，因此，对所有的$x$，

$$(g-f)(x) = \sum_n (g_n-f_n)(x).$$

令$h_n := g-f-\sum_{1\leqslant j\leqslant n}(g_j-f_j)$，那么$h_n\in\mathcal{L}$且$h_n\downarrow 0$，所以$I(h_n)\downarrow 0$. 又因为$I(g-f)=\sum_{j\geqslant 1}I(g_j-f_j)$，所以，$\nu$在$\mathcal{S}$上是可数可加的．□

ν在σ-代数上的扩张将仍记为ν. 斯通对这个理论的主要贡献是以X上关于测度μ的积分$\int f\mathrm{d}\mu$表示$I(f)$. 为定义这样一个测度μ，斯通发现可加性假设是很有用的．向量格$\mathcal{L}$称为斯通向量格，当且仅当对所有的$f\in\mathcal{L}$，$f\wedge 1\in\mathcal{L}$.（一般地，常函数不属于$\mathcal{L}$，任意包含常函数的向量格是一个斯通向量格．）

例：设$X:=[0, 1]$，$f(x)\equiv x$，$\mathcal{L}=\{cf: c\in\mathbb{R}\}$，$I(cf)=c$，那么$\mathcal{L}$是一个向量格，但不是一个斯通向量格，且$I$是$\mathcal{L}$上的一个准整数．

然而，实际上，具有实积分的向量格也是适用的，比如连续、可积或$\mathcal{L}^p$函数，仍然是斯通向量格，所以满足斯通条件．

正如定理4.1.6.之前所定义的，对X的子集构成的σ-环$\mathcal{B}$，X上的实值函数f称为可测的，当且仅当对$\mathbb{R}$中的每个不包含0的博雷尔集A，有$f^{-1}(A)\in\mathcal{B}$. 如果$\mathcal{B}$是一个σ-代数，由于$f^{-1}\{0\}=X\setminus f^{-1}(\mathbb{R}\setminus\{0\})$，所以这等价于通常的定义．

下面给出本节的主要定理.

4.5.2 定理(Stone-Daniell)　*设I是斯通向量格$\mathcal{L}$上的一个准整数，那么存在X上的一个测度μ，使*

得对所有的$f\in\mathcal{L}$，$I(f)=\int f\mathrm{d}\mu$. 在最小σ-环$\mathcal{B}$上，测度μ是唯一确定的，$\mathcal{L}$中所有的函数在最小σ-环$\mathcal{B}$上是可测的．

144

证明 设$\mathcal{L}$是定理4.5.1中所定义的．通过外测度ν^*，定理3.1.4的证明给出了ν到$\mathcal{T}$上的可数可加测度的一个特殊扩张．尽管ν可能不是σ-有限的，但这也是我们通过ν在$\mathcal{T}$中找到的，所以ν在由$\mathcal{S}$上到$\mathcal{T}$上的测度可能存在另外一个扩张．

对$f\in\mathcal{L}$，设$\mathcal{M}$是所有集合$f^{-1}((1,\infty))$构成的集类，那么对所有的$f\in\mathcal{L}$和$r>0$，$\mathcal{M}$包含集合$f^{-1}((r,\infty))=(f/r)^{-1}((1,\infty))$和$f^{-1}((-\infty,-r))=(-f)^{-1}((r,\infty))$.

由于对$r>0$，区间$(-\infty,-r)$和(r,∞)生成了$\mathbb{R}\setminus\{0\}$的博雷尔子集的σ-环，定理4.1.6意味着$\mathcal{M}$生成了该定理中定义的σ-环$\mathcal{B}$.

令$f\geqslant 0$，$f\in\mathcal{L}$，$g_n:=(n(f-f\wedge 1))\wedge 1$，那么对任意的$c>0$，$[0,cg_n)\uparrow f^{-1}((1,\infty))\times[0,c)$. 因此，对每个$A\in\mathcal{M}$，有$A\times[0,c)\in\mathcal{T}$. 对每个$A\in\mathcal{B}$，令$\mu(A):=\nu(A\times[0,1))$，由于$\{A: A\times[0,1)\in\mathcal{T}\}$是$\sigma$-环，因此$\mu$是明确定义的．

下面证明对任意的$A\in\mathcal{B}$和$c>0$，$\nu(A\times[0,c))=c\nu(A\times[0,1))$. 对$x\in X$，$t\in\mathbb{R}$，和$0<c<\infty$，令$M_c((x,t)):=(x,ct)$，那么$M_c$是从$X\times\mathbb{R}$映射到其自身的一一对应．对任意的$E\subset X\times\mathbb{R}$，令$M_c[E]:=\{M_c((x,t)):(x,t)\in E\}$. 注意到，$M_c\equiv M_{1/c}^{-1}$保持了所有的集合运算．对$\mathcal{L}$中任意的$f\leqslant g$，有$D:=[f,g)\in\mathcal{S}$，显然$M_c[D]=[cf,cg)\in\mathcal{S}$且$\nu(M_c[D])=c\nu(D)$. 因此，$E\in\mathcal{T}$当且仅当$M_c[E]\in\mathcal{T}$. 因为$\mathcal{S}$是一个环，根据命题3.2.3和命题3.2.4，对由$\mathcal{S}$生成的环$\mathcal{R}$及对所有的$E\in\mathcal{R}$，有$\nu(M_c[E])=c\nu(E)$. 根据外测度的定义，对所有的$E\subset X\times\mathbb{R}$，$\nu^*(M_c[E])=c\nu^*(E)$. 因此，对所有的$E\in\mathcal{T}$，$\nu(M_c[E])\equiv c\nu(E)$.

对集合A，令$E=[0,1_A)$，则对任意的$c>0$，有$M_c[E]=A\times[0,c)$. 从而对所有的$A\in\mathcal{B}$，$\nu(A\times[0,c))=c\nu(A\times[0,1))=c\mu(A)$.

正如由命题4.1.5给出的，设f_k是关于$\mathcal{B}$的简单函数，且$0\leqslant f_k\uparrow f$，那么$[0,f_k)\uparrow[0,f)$. 对任意关于$\mathcal{B}$的简单函数$h\geqslant 0$，某个$c_i>0$和某些不相交集$A(i)\in\mathcal{B}$，有$h=\sum_i c_i 1_{A(i)}$. 因此，$[0,h)$是集合$[0,c_i1_{A(i)})$的不交并，$\nu([0,h))=\sum_i\nu([0,c_i1_{A(i)}))=\sum_i c_i\mu(A(i))=\int f\mathrm{d}\mu$. 对每个$k$，对$h=f_k$运用上面的结论，再令$k\to\infty$，运用单调收敛性可得，

$$I(f)=\nu([0,f))=\lim_{k\to\infty}\nu([0,f_k))=\lim_{k\to\infty}\int f_k\mathrm{d}\mu=\int f\mathrm{d}\mu.$$

那么对一般的$f\in\mathcal{L}$，令$f=f^+-f^-$，对$g=f^+$和f^-运用$I(g)=\int g\mathrm{d}\mu$可证得对$g=f$，上面的式子成立．

下面证明唯一性，令ε是由所有集合A构成的集类，A在$\mathcal{B}$中且对任意两个测度μ和γ，以及所有的$f\in\mathcal{L}$，有$I(f)=\int f\mathrm{d}\mu=\int f\mathrm{d}\gamma$，从而$\mu(A)=\gamma(A)$. 在前面的证明中，对$A\in\mathcal{M}$，有$0\leqslant g_n\in\mathcal{L}$且$g_n\uparrow 1_A$. 因此，$\mathcal{M}\subset\varepsilon$. 显然对每个$A\in\mathcal{M}$，$\mu(A)<\infty$. 从而，$f^{-1}((1,\infty))\cup g^{-1}((1,\infty))=(f\vee g)^{-1}((1,\infty))$且$f^{-1}((1,\infty))\cap g^{-1}((1,\infty))=(f\wedge g)^{-1}((1,\infty))$. 因此，如果$C\in\mathcal{M}$和$D\in\mathcal{M}$，则有$C\cup D\in\mathcal{M}$和$C\cap D\in\mathcal{M}$. 又由于$1_{C\setminus D}=1_C-1_{C\cap D}$，从而$C\setminus D\in\varepsilon$. 根据命题3.2.8，由$\mathcal{M}$生成的环中的每个集合都是有限的，对$\mathcal{M}$中的$C_i$和$D_i$，它是$C_i\setminus D_i$的不交并，所以这个环包含在$\varepsilon$中．

145

现在，$\mathcal{B}$中每个集合都包含在$\mathcal{M}$中集合的可数并中（由于所有满足这一条件的集合构成的集类是一个 σ-环），又因为$\mathcal{M}$中的集合有有限测度（对 μ 和 γ），正如在定理 3.1.10 那样得，在$\mathcal{B}$上，$\mu=\gamma$. □

习题

1. 设 f 是集合 X 上的函数，$\mathcal{L}:=\{cf: c\in\mathbb{R}\}$，
   (a) 证明：如果 $f\geqslant 0$，$\mathcal{L}$是一个向量格；如果 $f\leqslant 0$，$\mathcal{L}$则不是一个向量格.
   (b) f 满足什么条件时，$\mathcal{L}$是斯通向量格？
   (c) 如果 I 在$\mathcal{L}$上是一个准整数，对某个测度 μ，$I(f)=\int f\mathrm{d}\mu$ 是否总是成立的？
   (d) 如果(c)中的 μ 存在，f 满足什么条件时，μ 是唯一的？
2. 对 $k=1,\cdots,n$，令 X_k 是不相交的集合，$\mathcal{F}_k$是 X_k 上的函数的向量格，设 $X=\bigcup_{1\leqslant k\leqslant n}X_k$，是 X 上所有函数 $f=(f_1,\cdots,f_n)$ 的集合，满足对每个 $k=1,\cdots,n$，$f_k\in\mathcal{F}_k$且在 X_k 上，$f=f_k$，
   (a) 证明：$\mathcal{F}$是一个向量格.
   (b) 如果对每个 k，I_k 是$\mathcal{F}_k$上的准整数. 对 $f\in\mathcal{F}$，令 $I(f):=\sum_{1\leqslant k\leqslant n}I_k(f_k)$，证明：$I$ 是$\mathcal{F}$上的一个准整数.（从而 $(X,\mathcal{F},I)$ 称为 $(X_k,\mathcal{F}_k,I_k)$ 的直和.）
3. 设 I 是向量格$\mathcal{L}$上的准整数，$\mathcal{U}$是所有函数 f 组成的集合，使得对所有的 n，$f_n\in\mathcal{L}$，且对所有的 x，$f_n(x)\uparrow f(x)$. 令 $\bar{I}(f):=\lim_{n\to\infty}I(f_n)\leqslant\infty$，证明：$\bar{I}$ 在$\mathcal{U}$上是明确定义的.
4. 设 $I^{\mathbb{N}}$ 是所有序列 $\{x_n\}$ 的集合，其中对所有的 n，$x_n\in I:=[0,1]$，所以 $I^{\mathbb{N}}$ 是一个无限维的单位立方体，问题是在 $I^{\mathbb{N}}$ 上构造一个无限维的勒贝格测度. 设$\mathcal{L}$是 $I^{\mathbb{N}}$ 上所有实值函数的集合，使得对某个有限的 n 和 I^n 上的连续函数 g，有 $f(\{x_j\}_{j\geqslant 1})\equiv g(\{x_j\}_{1\leqslant j\leqslant n})$.
   (a) 证明：$\mathcal{L}$是一个斯通向量格.
   (b) 对如上的 f 和 g，令 $I(f):=\int_0^1\cdots\int_0^1 g(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n$，证明：$I$ 是一个准整数.

146

5. 设$\mathcal{L}$是满足当 $n\to\infty$ 时，x_n 收敛的所有实数序列 $\{x_n\}_{n\geqslant 1}$ 的集合，
   (a) 证明：$\mathcal{L}$是一个斯通向量格.
   (b) 在$\mathcal{L}$上，令 $I(\{x_n\}_{n\geqslant 1})=\lim_{n\to\infty}x_n$，证明：$I$ 不是一个准整数.
   (c) 令$\mathcal{M}$是满足当 $n\to\infty$ 时，$x_n\to x_0$ 的所有序列 $\{x_n\}_{n\geqslant 0}$ 的集合，证明：$\mathcal{M}$是一个斯通向量格，如(b)中所定义的，I 是一个准整数.
6. 设$\mathcal{P}_2$是$\mathbb{R}$上次数不超过 2 的所有多项式 Q 的集合，所以对某些实数 a，b 和 c，$Q(x)\equiv ax^2+bx+c$. 对这样的 Q，令 $I(Q):=a$. 证明：
   (a) I 在不是格的向量空间$\mathcal{P}_2$上是线性的.
   (b) 对任意的 $f\geqslant 0$，$f\in\mathcal{P}_2$，$I(f)\geqslant 0$.
   (c) 对任意的 $f_n\downarrow 0$，$f\in\mathcal{P}_2$，$I(f_n)\downarrow 0$.
   (d) 证明不存在包含$\mathcal{P}_2$的向量格$\mathcal{L}$，使得 I 可以扩张为$\mathcal{L}$上的一个准整数.[提示：对 g 有界，$g\in\mathcal{L}$，有 $I(g)=0$.]
7. 如果在定理 4.5.2 中 μ 是 σ-有限的，证明：在$\mathcal{T}$上，$\nu=\mu\times\lambda$.
8. 给出一个例子，其中斯通一丹尼尔定理中的测度 μ 到最小 σ-代数$\mathcal{A}$的扩张不止一个，$\mathcal{L}$中的所有函数对$\mathcal{A}$是可测的，而且在最小 σ-环$\mathcal{R}$上，μ 是有界的且对 σ-环$\mathcal{R}$函数是可测的.[提示：参见 4.1 节习题 9.]
9. 给出一个类似的例子，但 μ 在$\mathcal{R}$上是无界的.
10. 证明：在上面两个习题的条件下，总存在 μ 的一个最小扩张 ν，即对 μ 到$\mathcal{A}$上测度的任意扩张 ρ，使得对所有的 $B\in\mathcal{A}$，$\nu(B)\leqslant\rho(B)$ 成立.

11. 设 g_1，…，g_k 是集合 X 上的线性无关的实值函数，即如果对实常数 c_j，$\sum_{j=1}^{k} c_j g_j \equiv 0$，则 $c_1 = c_2 = \cdots = 0$. 证明：对 X 中的某些 x_1，…，x_k，g_1，…，g_j 在 $\{x_1, \cdots, x_j\}$ 上是线性无关的，其中 $j = 1$，…，k.
12. 设$\mathcal{F}$是集合 X 上的实值函数的一个有限维向量空间．假设对所有的 $x \in X$，$f_n \in \mathcal{F}$且 $f_n(x) \to f(x)$．证明$f \in \mathcal{F}$. [提示：利用习题 11.]
13. 设$\mathcal{F}$是由三个点组成的集合 $\{p, q, r\}$ 上的一个函数向量格，满足对某些 b 和 c，以及$\mathcal{F}$中所有的 f，有 $f(p) \equiv bf(q) + cf(r)$．证明：或者$\mathcal{F}$是一维的，如习题 1 中，或者 $b \geqslant 0$，$c \geqslant 0$ 且 $bc = 0$.
14. 设$\mathcal{F}$是集合 X 上的一个函数向量格，且$\mathcal{F}$是一个有限维向量空间．证明：$\mathcal{F}$是一个一维向量格(同习题 1(a))的直和(同习题 2)．[提示：利用习题 11 和 13 的结果．]

147

15. 如果$\mathcal{F}$是一个有限维斯通向量格，证明：在习题 14 表示的直和中，对某些集合 $X(i)$ 及每个 i，有 $\mathcal{F}_i = \{c1_{X(i)}: c \in \mathbb{R}\}$.
16. 如果 I 是一个有限维向量格$\mathcal{F}$上的非负线性函数，证明：I 是一个准整数，并且对某个有限测度 μ，对所有的$f \in \mathcal{F}$，有 $I(f) = \int f d\mu$. 证明：在最小 σ-环上 μ 是唯一确定的，对这个最小 σ-环$\mathcal{F}$中的所有函数是可测的，当且仅当$\mathcal{F}$是一个斯通向量格．[提示：利用习题 14 和 15 的结果．]
17. 如果 V 是一个实值函数的向量空间，使得对某个 $n < \infty$ 和所有的$f \in V$，f的值域的势至多为 n，证明：V 的维数至多为 n.

注释

4.1 节 现代的积分概念归功于 Henri Lebesgue(1902). Hawkins(1970)介绍了这个概念的发展历史，Medvedev(1975)收录了勒贝格个人的著作．勒贝格积分的发现是分析中的一个重大突破．勒贝格的著作集已发表五卷书(Lebsgue，1972—1973)，前两卷包括了他关于"积分和微分"的工作．第一卷还包括了三篇文章，这三篇文章包含了勒贝格的生活及其与 Arnaud Denjoy-Lucienne Felix 和 Paul Montel 等人的相关工作，同时也介绍了勒贝格自己的工作，并用 80 页介绍了到 1922 年其工作的接受度．May(1966)也给出了勒贝格的一个简要传记，他生于 1875 年卒于 1941 年．

像测度定理 4.1.11 在欧几里得空间中可以采取一种不同的形式来描述(参考 4.4 节的注释).

4.2 节 命题 4.2.1 是在 Halmos(1950，p.83)中以练习的形式出现的．他没有给出更早的参考文献．Hausdorff(1914，p.390—392)证明了定理 4.2.2. 命题 4.2.3 出现在 Dudley(1971，命题 3)中．定理 4.2.5 归功于 von Alexits(1930)和 Sierpiński(1930)．这里给出的证明基本上与 Lehmann(1959，p.37—38)类似．拓展到完备可分值域空间(这里是通过命题 4.2.6 完成的)归功于 Kuratowski(1933；1966，p.434)．Shortt(1983)考虑了更一般的值域空间．定理 4.2.8 出现在 Lehmann(1959，p.37，引理 1)中，Lehmann 给出了一些更早的参考文献，但我不知道它首次出现是什么时候．非常感谢 Deborah Allinger 和为这些注释提供信息的 Rae Shortt.

4.3 节 在有界区间或$\mathbb{R}$ 的可测子集上关于勒贝格测度的所有极限定理首次得到了证明．在这种情况下：Beppo Levi(1906)证明了单调收敛定理 4.3.2. Fatou(1906)提出了引理 4.3.3. 根据 Nathan(1971)，尽管 Fatou(1878—1929)学的是数学专业，是一位有名的数学家，但是他也是一位天文学家，一直都在巴黎观察台工作．Vitali(1907)证明了一致可积序列的收敛定理(见定理 10.3.6)，其中包括控制收敛定理 4.3.5. Lebesgue(1902，§25)证明了一致有界函数序列的控制收敛性，但仍然是在有界区间上．

148

Lebesgue(1910，p.375)详细地阐明了控制收敛定理，如今是关于多维勒贝格测度 λ^k 的．

4.4 节 Lebesgue(1902，p.37—390)在 Lebesgue(1972—1973，Vol.2)证明了对有限的 a，b，c，

d 和任意有界可测函数 $f(x, y)$ 积分 $\int_a^b dx$ 和 $\int_c^d dy$ 是可以交换的．根据 Saks(1937, p. 77)，由"柯西准则"，f 是连续的．$\mathbb{R}^2$ 中的可测集上的示性函数 f 给出了关于有界区间上勒贝格测度的乘积测度存在定理(4.4.4)．Lebesgue(1902，绪论)阐述了将其扩张为两个以上的变量(如在定理 4.4.6 中?)是可行的．遗憾的是，Lebesgue(1902，§40)继续断言当 f 不一定必须有界也不可测时，积分仍然可交换，如果所有的积分存在．这对例子中的可测无界函数是不成立的，也是因为 Cauchy(Hawkins, 1970, p. 91)而出名，此例子中的可测无界函数与本书中的(a)和(b)类似，且对有界不可测函数(习题 4)也不成立．

Fubini(1907)阐述了关于对勒贝格测度交换积分次序的定理(如 4.4.5)，并且定理 4.4.5 一般称为"Fubini 定理"．但是富比尼的证明是有"缺陷的"(Hawkins, 1970, p. 161)．显然，Tonelli(1909)给出了第一个正确的证明，这一证明并入为富比尼定理的一部分．

采用连续统假设，将习题 4 应用于勒贝格测度．H. Friedman(1980)证明了这与通常的(Zermelo-Fraenkel，见附录 A)包含选择公理的集合论是一致的，还证明了在任何时候一个在 $[0, 1]\times[0, 1]$ 上有界(可能不可测)的函数 f 都是使得累积分 $\iint f dx dy$ 和 $\iint f dy dx$ 有定义且相等的函数．所以连续统假设不能被省略．在这一假设下，存在相当特殊的集合：Sierpiński(1920)，借助无限递归式，在 $[0, 1]\times[0, 1]$ 上定义一个集合 A，且集合内每条线(不仅仅是那些与轴平行的线)至多有两个交点，但对于 $\lambda\times\lambda$ 外测度为 1. Bledsoe-Morse(1955)定义了扩张的乘积测度(习题 14)，它给出了谢尔品斯基(Sierpiński)集的测度为 0. 采用连续统假设，由习题 4 中的例子，Bledsoe-Morse 乘积测度不可避免地违背了 Friedman 的阐述．

对 $\mathbb{R}^k$ 上的勒贝格测度 λ^k，像测度定理 4.1.11 有一个更具体的形式，其中 T 是从 $\mathbb{R}^k$ 映射到它自身的 1－1 函数，T 的逆 T^{-1} 有连续的一阶偏导数：在某些条件下，$\lambda^k\circ T^{-1}$ 可被 T^{-1} 的 λ^k 次雅可比行列式所代替．例如，参见 Rudin(1976, p. 252)．

4.5 节　Daniell(1917—1918)发展了他的积分理论，对有界函数空间，从准整数 I 开始，并将它扩张到一个大的函数类 $\mathcal{L}^1(I)$ 上，类似的且在许多情况下与关于测度的积分理论相一致．稍后，Stone(1948)证明了在对每个 $f\in\mathcal{L}$，$f\wedge 1\in\mathcal{L}$ 的条件下，准整数 I 可以用关于测度的积分来表示．另外，丹尼尔由于对数学统计学所做的一些贡献而闻名(参见 Stigler, 1973)．A. C. Zaanen(1958, 13 节)证明了定理 4.5.1，并把它作为他对斯通–丹尼尔定理(4.5.2)的新的、简短证明的一部分．149

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到．

Alexits, G. von (1930). Über die Erweiterung einer Baireschen Funktion. *Fund. Math.* *15*: 51–56.

Bledsoe, W. W., and Anthony P. Morse (1955). Product measures. *Trans. Amer. Math. Soc.* *79*: 173–215.

Daniell, Percy J. (1917–1918). A general form of integral. *Ann. Math.* *19*: 279–294.

Dudley, R. M. (1971). On measurability over product spaces. *Bull. Amer. Math. Soc.* *77*: 271–274.

Fatou, Pierre Joseph Louis (1906). Séries trigonométriques et séries de Taylor. *Acta Math.* *30*: 335–400.

Friedman, Harvey (1980). A consistent Fubini-Tonelli theorem for nonmeasurable functions. *Illinois J. Math.* *24*: 390–395.

*Fubini, Guido (1907). Sugli integrali multipli. *Rendiconti Accad. Nazionale dei Lincei* (Rome) (Ser. 5) *16*: 608–614.

Halmos, Paul R. (1950). *Measure Theory*. Van Nostrand, Princeton. Repr. Springer, New York (1974).

Hausdorff, Felix (1914) (see references to Chapter 1).

Hawkins, T. (1970). *Lebesgue's Theory of Integration, Its Origins and Development*. University of Wisconsin Press.

Hörmander, Lars (1983). *The Analysis of Linear Partial Differential Operators*, I. Springer, Berlin.

Kuratowski, C. [Kazimierz] (1933). Sur les théorèmes topologiques de la théorie des fonctions de variables réelles. *Comptes Rendus Acad. Sci. Paris 197*: 19–20.

——— (1966). *Topology*, vol. 1. Academic Press, New York.

Lebesgue, Henri (1902). Intégrale, longueur, aire (thèse, Univ. Paris). *Annali Mat. pura e appl.* (Ser. 3) *7*: 231–359. Also in Lebesgue (1972–1973) *2*, pp. 11–154.

——— (1904). *Leçons sur l'intégration et la recherche des fonctions primitives*. Paris.

——— (1910). Sur l'intégration des fonctions discontinues. *Ann. Scient. Ecole Normale Sup*. (Ser. 3) *27*: 361–450. Also in Lebesgue (1972–1973), *2*, pp. 185–274.

——— (1972–1973). *Oeuvres scientifiques*. 5 vols. L'Enseignement Math., Inst. Math., Univ. Genève.

Lehmann, Erich (1959). *Testing Statistical Hypotheses*. 2d ed. Wiley, New York (1986).

Levi, Beppo (1906). Sopra l'integrazione delle serie. *Rend. Istituto Lombardo di Sci. e Lett.* (Ser. 2) *39*: 775–780.

May, Kenneth O. (1966). Biographical sketch of Henri Lebesgue. In Lebesgue, H., *Measure and the Integral*, Ed. K. O. May, Holden-Day, San Francisco, transl. from *La Mesure des grandeurs, Enseignement Math 31–34* (1933–1936), repub. as a *Monographie* (1956).

Medvedev, F. A. (1975). The work of Henri Lebesgue on the theory of functions (on the occasion of his centenary). *Russian Math. Surveys 30*, no. 4: 179–191. Transl. from *Uspekhi Mat. Nauk 30*, no. 4: 227–238.

Nathan, Henry (1971). Fatou, Pierre Joseph Louis. In *Dictionary of Scientific Biography, 4*, pp. 547–548. Scribner's, New York.

Rudin, Walter (1976). *Principles of Mathematical Analysis*. 3d ed. McGraw-Hill, New York.

Saks, Stanisław (1937). *Theory of the Integral*. 2d ed. English transl. by L. C. Young. Hafner, New York. Repr. corrected Dover, New York (1964).

Schwartz, Laurent (1966). *Théorie des Distributions*. 2d ed. Hermann, Paris.

Shortt, R. M. (1983). The extension of measurable functions. *Proc. Amer. Math. Soc. 87*: 444–446.

Sierpiński, Wacław (1920). Sur un problème concernant les ensembles measurables superficiellement. *Fundamenta Math. 1*: 112–115.

——— (1930). Sur l'extension des fonctions de Baire définies sur les ensembles linéaires quelconques. *Fund. Math. 16*: 81–89.

Stigler, Stephen M. (1973). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *J. Amer. Statist. Assoc. 68*: 872–879.

Stone, Marshall Harvey (1948). Notes on integration, II. *Proc. Nat. Acad. Sci. USA 34*: 447–455.

Tonelli, Leonida (1909). Sull'integrazione per parti. *Rendiconti Accad. Nazionale dei Lincei* (Ser. 5) *18*: 246–253. Reprinted in Tonelli, L., *Opere Scelte* (1960), *1*, pp. 156–165. Edizioni Cremonese, Rome.

*Vitali, Giuseppe (1904–05). Sulle funzioni integrali. *Atti Accad. Sci. Torino 40*: 1021–1034.

——— (1907). Sull'integrazione per serie. *Rend. Circolo Mat. Palermo 23*: 137–155.

Zaanen, Adriaan C. (1958). *Introduction to the Theory of Integration*. North-Holland, Amsterdam.

第5章 L^p 空间：泛函分析引论

泛函分析的中心思想是把函数视为函数空间中的“点”．首先，考虑有穷测度空间$(X, \mathcal{S}, \mu)$（如具有勒贝格测度的单位区间）上的有界可测函数．对任意的两个函数f和g，我们可得到一个有限积分$\int fg\mathrm{d}\mu = \int f(x)g(x)\,\mathrm{d}\mu(x)$．如果把函数视为向量，则此积分有内积或点积$(f, g)$的性质：当$f=g$时，$(f, g)$是非负的；在$(f, g)=(g, f)$的意义下$(f, g)$是对称的，并且对固定的$g$，关于$f$是满足线性性的．正如在有限维向量空间中那样，利用此内积，在函数空间上可以得到一个类似的欧几里得几何，距离为$d(f, g)=(f-g, f-g)^{1/2}$．实际上，如果μ是k元有限点集上的计数测度，则(f, g)变为了$\mathbb{R}^k$中通常向量的内积．但是，例如，如果μ是$[0, 1]$上的勒贝格测度，为了使得具有距离d的函数度量空间是完备的，我们就需要一些使得$\int f^2\mathrm{d}\mu < \infty$成立的无界函数$f$．根据同样的思路，对任意的$p>0$和$\mu$，存在函数$f$的集类，其中函数$f$是可测的且满足$\int |f|^p\mathrm{d}\mu < \infty$，这个集类称为$\mathcal{L}^p$或$\mathcal{L}^p(\mu)$．如果$f$和$g$属于$\mathcal{L}^p$，是否也有$f+g$属于$\mathcal{L}^p$，这并不显然．与此相应，对两个独立的函数$f$，$g$在相应积分形式下，是否有不等式$\int |f+g|^p\mathrm{d}\mu$成立，这些问题将在5.1节中讨论．在5.3节和5.4节中，将要进一步介绍$p=2$时的内积．

最后以一个重要的事实来完成测度论的讨论，即关于两个测度的关系的“Radon-Nikodym”定理，这将在5.5节中用一些泛函分析的知识来给予证明．

5.1 积分不等式

在具有勒贝格测度的$(0, 1]$上，函数$f(x)=x^p$是可积的当且仅当$p>-1$．然而两个可积函数的乘积未必可积，例如，$x^{-1/2}\cdot x^{-1/2}$不可积．因此，下面给出乘积可积的条件．

152

定义 对任意的测度空间$(X, \mathcal{S}, \mu)$和$0<p<\infty$，$\mathcal{L}^p(X, \mathcal{S}, \mu) := \mathcal{L}^p(X, \mathcal{S}, \mu, \mathbb{R})$表示$X$上的所有可测函数$f$组成的集合，其中$f$满足$\int |f|^p\mathrm{d}\mu < \infty$且除了在一个零测集上外，$f$的所有取值都为实数．这里$f$可能无定义或是无限的．对$1\leqslant p<\infty$，令$\|f\|_p = \left(\int |f|^p\mathrm{d}\mu\right)^{1/p}$，这称为$f$的“$L^p$范数”或“$p$-范数”．

设$\mathbb{C}$为复平面，如附录B中所定义的．$\mathbb{C}$上的一个函数f可写成$f=g+\mathrm{i}h$，其中g，h是实值函数，分别称为f的实部和虚部．$\mathbb{C}$将具有$\mathbb{R}^2$上的通常拓扑，所以f是可测的当且仅当g，h是可测的（见前面命题4.1.7的讨论）．空间$\mathcal{L}^p(X, \mathcal{L}, \mu, \mathbb{C})$称为复$\mathcal{L}^p$，与其相反，上段中的空间称为“实$\mathcal{L}^p$”，定义为$X$上的所有可测函数$f$组成的集合，其中$f$满足$\int |f|^p\mathrm{d}\mu<\infty$．且除了一个零测集外，$f$取值于$\mathbb{C}$，在此零测集上$f$可能无定义，$p$-范数和前面的定义一样．

下面是第一基本不等式．

5.1.1 定理 对任意取值于$\mathbb{R}$或$\mathbb{C}$上的可积函数f，有$\left|\int f\mathrm{d}\mu\right| \leqslant \int |f|\,\mathrm{d}\mu$．

证明 首先，如果f为实值函数，

$$\left|\int f\mathrm{d}\mu\right| = \left|\int f^+\mathrm{d}\mu - \int f^-\mathrm{d}\mu\right| \leqslant \left|\int f^+\mathrm{d}\mu + \int f^-\mathrm{d}\mu\right|$$
$$= \int f^+ + f^-\mathrm{d}\mu = \int |f|\,\mathrm{d}\mu.$$

现在，如果f为复值函数．设$f=g+\mathrm{i}h$，其中g，h为实值函数．那么只需证

$$\left\{\left(\int g\mathrm{d}\mu\right)^2 + \left(\int h\mathrm{d}\mu\right)^2\right\}^{1/2} \leqslant \int (g^2+h^2)^{1/2}\mathrm{d}\mu.$$

在$\mathbb{R}^2$中将坐标旋转θ角，以$(g\cos\theta - h\sin\theta, g\sin\theta + h\cos\theta)$代替$(g, h)$，不等式两边仍成立．所以，沿正$x$轴将向量$\left(\int g\mathrm{d}\mu, \int h\mathrm{d}\mu\right)$旋转到一点，我们可以假设$\int h\mathrm{d}\mu = 0$. 那么从取实值情况时可以得出，$\left|\int g\mathrm{d}\mu\right| \leqslant \int |g|\,\mathrm{d}\mu$，并且$|g| \leqslant (g^2+h^2)^{1/2}$，定理证毕(由引理4.1.9). □

定理5.1.1中的不等式变为等式，当且仅当f是非负的或更一般地，$f\equiv cg$，这里$g\geqslant 0$，c是一个固定的复数．

153

第二个不等式可能有些出乎意料，除去$p=2$的情况(推论5.1.4)外．这个不等式将用在5.1.5的基本不等式($\|\cdot\|_p$的次可加性)的证明中．

5.1.2 定理(Rogers-Hölder 不等式) 如果$1<p<\infty$，$p^{-1}+q^{-1}=1$，$f\in\mathcal{L}^p(X, \mathcal{S}, \mu)$和$g\in\mathcal{L}^q(X, \mathcal{S}, \mu)$，则$fg\in\mathcal{L}^1(X, \mathcal{S}, \mu)$且$|\int fg\mathrm{d}\mu| \leqslant \int |fg|\,\mathrm{d}\mu \leqslant \|f\|_p\|g\|_q$.

证明 如果$\|f\|_p=0$，则$f=0$a. e.，所以$fg=0$a. e.，$|\int fg\mathrm{d}\mu|=0$，不等式成立．同样，如果$\|g\|_q=0$，则不等式也成立．所以我们假定这两个范数都不为零，对任意常数$c>0$，$\|cf\|_p=c\|f\|_p$. 两端除以这个范数，可得$\|f\|_p=\|g\|_q=1$. 注意，fg是可测的(正如在前面命题4.1.8中证明的)．现在要用到下面的引理.

5.1.3 引理 对任意的正实数u和v，及$0<\alpha<1$，有$u^\alpha v^{1-\alpha}\leqslant \alpha u+(1-\alpha)v$.

证明 两边除以v，假设$v=1$. 即要证$u^\alpha\leqslant\alpha u+1-\alpha$，显然$u=1$时成立．现在两边对$u$求导，$u<1$时，$\alpha u^{\alpha-1}>\alpha$；$u>1$时，$\alpha u^{\alpha-1}<\alpha$.

通过微积分中值定理，这些事实蕴涵着这个引理． □

下面继续证明定理5.1.2，令$\alpha=1/p$，对每个x，令$u(x):=|f(x)|^p$，$v(x):=|g(x)|^q$，则$1-\alpha=1/q$，且由引理知，对所有的x，$|f(x)g(x)|\leqslant\alpha|f(x)|^p+(1-\alpha)|g(x)|^q$. 两边积分得$\int|fg|\,\mathrm{d}\mu\leqslant\alpha+(1-\alpha)=1$，于是证明了定理5.1.2中的主要(第二个)不等式，也就证明了$fg\in\mathcal{L}^1(X, \mathcal{S}, \mu)$. 由定理5.1.1得到第一个不等式，从而证明了定理5.1.2.

最熟悉最常用的是$p=2$的情况．

5.1.4 推论(Cauchy-Bunyakovsky-Schwarz 不等式) 对任意f，$g\in\mathcal{L}^2(X, \mathcal{S}, \mu)$，有$fg\in\mathcal{L}^1(X, \mathcal{S}, \mu)$且$\left|\int fg\mathrm{d}\mu\right|\leqslant\|f\|_2\|g\|_2$.

取 Rogers-Hölder 不等式的例子，令$X=[0, 1]$，其上的测度μ为勒贝格测度．对某个$r>0$，$s>0$，令$f(x)=x^{-r}$，$g(x)=x^{-s}$，则$fg\in\mathcal{L}^1$当且仅当$r+s<1$. 在边界情况$r+s=1$时，如在定理5.1.2中那样，取$p=1/r$，及$p^{-1}+q^{-1}=1$，所以$q=1/s$，则$f\notin\mathcal{L}^p$，但对$\forall t<p$，$f\in\mathcal{L}^t$，且$g\notin\mathcal{L}^q$，但$\forall u<q$，有$g\in\mathcal{L}^u$. 所以定理5.1.2的假设都不成立，但在边界的情况(正如以上结论)都是正

154

确的．这表明对假设条件的类型，定理 5.1.2 很难得到改进．定理 6.4.1 将给出一个更精确的描述．

$\mathcal{L}^p$空间对 $p=+\infty$ 也有定义．可测函数称为是本性有界的(essentially bounded)当且仅当 $M<\infty$，$|f|<M$a. e.．所有本性有界实值函数的集合称为$\mathcal{L}^\infty(X,\mathcal{S},\mu)$或$\mathcal{L}^\infty$，如果测度空间的元素是清楚的．同样，定义$\mathcal{L}^\infty(X,\mathcal{S},\mu,\mathbb{C})$为 X 上有定义、可测、本性有界并且几乎处处有复值的函数的集合．对任意的 f，令 $\|f\|_\infty:=\inf\{|f|\leqslant M\text{a. e.}\}$．所以对某些 $M_n\downarrow\|f\|_\infty$，除了在满足 $\mu(A_n)=0$ 的集合 A_n 上外，$|f|\leqslant M_n$. 因此，除了在 A_n 的并上外，对任意的 n，几乎处处有 $|f|\leqslant M_n$，所以 $|f|\leqslant\|f\|_\infty$a. e.．

如果 f 是$[0,1]$上的连续函数，μ 是勒贝格测度，则 $\|f\|_\infty=\sup|f|$，因为对 f 的任意值，它在正测度集上有相近的取值．

如果 $f\in\mathcal{L}^1$，$g\in\mathcal{L}^\infty$，则 $|fg|\leqslant|f|\|g\|_\infty$a. e.，所以 $fg\in\mathcal{L}^1$，$\int|fg|\,\mathrm{d}\mu\leqslant\|f\|_1\|g\|_\infty$. 因此，Rogers-Hölder 不等式扩展到了 $p=1$，$q=\infty$ 的情形．对 $p\geqslant1$，$\mathcal{L}^p$上的伪度量将通过 $d(f,g):=\|f-g\|_p$ 来定义．为了表明 d 满足三角不等式，我们需要下列事实．

5.1.5 定理(Minkowski-Riesz 不等式)　对 $1\leqslant p\leqslant\infty$，如果 f，$g\in\mathcal{L}^p(X,\mathcal{S},\mu)$，则

$$f+g\in\mathcal{L}^p(X,\mathcal{S},\mu)\text{ 且 }\|f+g\|_p\leqslant\|f\|_p+\|g\|_p.$$

证明　首先，如同命题 4.1.8 中那样，$f+g$ 是可测的．因为 $|f+g|\leqslant|f|+|g|$，所以可以用 f 和 g 的绝对值代替 f 和 g，于是假设 $f\geqslant0$，$g\geqslant0$. 如果 $f=0$a. e. 或 $g=0$a. e.，不等式是显然成立的．对 $p=1$ 或∞，不等式可以直接得到．

对 $1<p<\infty$，有 $(f+g)^p\leqslant2^p\max(f^p,g^p)\leqslant2^p(f^p+g^p)$，所以 $f+g\in\mathcal{L}^p$. 则由 Rogers-Hölder 不等式(定理 5.1.2)得，

$$\begin{aligned}\|f+g\|_p^p&=\int(f+g)^p\mathrm{d}\mu=\int f(f+g)^{p-1}\mathrm{d}\mu+\int g(f+g)^{p-1}\mathrm{d}\mu\\&\leqslant\|f\|_p\|(f+g)^{p-1}\|_q+\|g\|_p\|(f+g)^{p-1}\|_q.\end{aligned}$$

而 $(p-1)q=p$，因此 $\|f+g\|_p^p\leqslant(\|f\|_p+\|g\|_p)\|f+g\|_p^{p/q}$. 又因为 $p-p/q=1$，所以在上述不等式两端同除以 $\|f+g\|_p^{p/q}$ 得，$\|f+g\|_p\leqslant\|f\|_p+\|g\|_p$. □

如果 X 是一个具有计数测度的有限集，且 $p=2$，则定理 5.1.5 可以归约为关于通常的欧几里得距离的三角不等式．

令 X 是一个实向量空间(如线性代数中所定义的或特别地，在附录 B 所定义的)．X 上的半范数(seminorm)是一个从 X 映射到$[0,\infty)$上的函数$\|\cdot\|$，满足

155

(i)对所有 $c\in\mathbb{R}$，$x\in X$，$\|cx\|=|c|\,\|x\|$.

(ii)对所有 x，$y\in X$，$\|x-y\|\leqslant\|x\|+\|y\|$.

一个半范数$\|\cdot\|$称为范数(norm)当且仅当 $\|x\|=0$ 仅对 $x=0$ 成立．(在任意情况下，$\|0\|=\|0\cdot0\|=0\cdot\|0\|=0$.)

例：Minkowski-Riesz 不等式(定理 5.1.5)蕴涵了对任意度量空间$(X,\mathcal{S},\mu)$及 $1\leqslant p\leqslant\infty$，$\mathcal{L}^p(X,\mathcal{S},\mu)$是一个向量空间且$\|\cdot\|_p$是其上的一个半范数．如果集合 A 非空且 $\mu(A)=0$，则对每个 p，$\|1_A\|_p=0$，但$\|\cdot\|_p$不是$\mathcal{L}^p$上的范数．对任意一个半范数$\|\cdot\|$，令 $d(x,y):=\|x-y\|$ 对任意的 x，$y\in X$ 成立．则 d 是一个伪度量，即除了对某些 $x\neq y$，$d(x,y)=0$ 外，它满足了度量的所有条件．所以 d 是一个度量当且仅当$\|\cdot\|$是一个范数．

下面的不等式有时也会用到.

5.1.6 算术-几何平均不等式 对任意的非负数 x_1, x_2, …, x_n, $(x_1x_2\cdots x_n)^{1/n}\leqslant(x_1+\cdots+x_n)/n$.

注：这里 $(x_1x_2\cdots x_n)^{1/n}$ 称为 x_1, x_2, …, x_n 的几何平均，$(x_1+\cdots+x_n)/n$ 称为 x_1, x_2, …, x_n 的算术平均.

证明 通过对 n 作归纳来证明. $n=1$ 不等式显然成立. 如果任意 x_i 为零，则不等式总是成立的，所以假设对所有 i, $x_i>0$. 对归纳步应用引理 5.1.3，令 $u=(x_1x_2\cdots x_{n-1})^{1/(n-1)}$, $v=x_n$, $\alpha=(n-1)/n$，所以 $1-\alpha=1/n$，不等式对 n 成立. □

习题

1. 证明或反证：对任意复数 z_1, …, z_n,
   (a) $|z_1z_2\cdots z_n|^{1/n}\leqslant(|z_1|+\cdots+|z_n|)/n$.
   (b) $|z_1z_2\cdots z_n|^{1/n}\leqslant|z_1+\cdots+z_n|/n$.
2. 给出定理 5.1.1 的另一种证明，不先证实值函数的情况，而是用 $\int f\mathrm{d}\mu=re^{i\theta}$ 的复极坐标分解，其中 $r\geqslant0$, θ 是实数且 $f=\rho e^{i\varphi}$，其中 $\rho\geqslant0$, φ 是实值函数.

156

3. 令 Y 是取值于 $\mathbb{R}^3$ 的函数，$Y=(f, g, h)$，其中 f, g 和 h 是关于测度 μ 的可积函数. 令 $\int Y\mathrm{d}\mu:=\left(\int f\mathrm{d}\mu,\int g\mathrm{d}\mu,\int h\mathrm{d}\mu\right)$，证明：$\left\|\int Y\mathrm{d}\mu\right\|\leqslant\int\|Y\|\mathrm{d}\mu$，如果 $\|(x, y, z)\|=$
   (a) $(x^2+y^2+z^2)^{1/2}$;　　(b) $|x|+|y|+|z|$;
   (c) $\max(|x|, |y|, |z|)$;　　(d) $(|x|^p+|y|^p+|z|^p)^{1/p}$，其中 $1<p<\infty$.
4. 令 $(Y, \|\cdot\|)$ 是一个线性空间，$\|\cdot\|$ 是一个半范数. 假设 Y 对于 $d(y, z):=\|y-z\|$ 是可分的. 令 $(X, \mathcal{S}, \mu)$ 是一个可测空间，$1\leqslant p<\infty$. 令 $\mathcal{L}^p(X, \mathcal{S}, \mu)$ 是从 X 映射到 Y 的函数 f 的集合，其中 f 对于由关于 d 的开集所生成的 Y 上的博雷尔 σ-代数是可测的，并且 $\int\|f\|^p\mathrm{d}\mu(X)<\infty$. 如果 $f\in\mathcal{L}^p(X, \mathcal{S}, \mu, Y)$ 和 $g\in\mathcal{L}^q(X, \mathcal{S}, \mu, \mathbb{R})$，其中 $p^{-1}+q^{-1}=1$, $1<p<\infty$，证明：$fg\in\mathcal{L}^1(X, \mathcal{S}, \mu, Y)$，且 $\|fg\|_1\leqslant\|g\|_q\|f\|_p$. [提示：对于可测性，证明 $(c, y)\mapsto cy$ 是联合连续：$\mathbb{R}\times Y\mapsto Y$ 从而由命题 4.1.7 知，它是联合可测的.]
5. 证明：对 $1\leqslant p<\infty$, $\|\cdot\|_p$ 是 $\mathcal{L}^p(X, \mathcal{S}, \mu, Y)$ 上的半范数.
6. 对 $1\leqslant p<\infty$, $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \mu)$ 是两个 σ-有限可测空间，并假设 $\mathcal{L}^p(Y, \mathcal{T}, \nu)$ 是可分的. 证明：$\mathcal{L}^p(X, \mathcal{S}, \mu)$, $\mathcal{L}^p(Y, \mathcal{T}, \nu, \mathbb{R})$ 和 $\mathcal{L}^p(X\times Y, \mathcal{S}\otimes\mathcal{T}, \mu\times\nu, \mathbb{R})$ 是等距同构的，即证明从一个空间到另一个空间有一个一一对应的函数，保持半范数 $\|\cdot\|_p$. [提示：在推论 4.2.7 中令 $U=X$，以得到对所有的 x，有 $\|g_n(x)\|\uparrow\|g(x)\|$.]
7. 对 $0<p<1$，令 $\mathcal{L}^p(X, \mathcal{S}, \mu)$ 是 X 上的所有实值可测函数，满足 $\int|f(x)|^p\mathrm{d}\mu(x)<\infty$.
   (a) 举例证明：积分的第 p 个根不一定是 $\mathcal{L}^p$ 上的半范数.
   (b) 证明：对 $0<p\leqslant1$, $d_p(f, g):=\int|f-g|^p\mathrm{d}\mu$ 定义了一个 $\mathcal{L}^p(X, \mathcal{S}, \mu)$ 上的伪度量.
8. 举例证明：对 $1<p<\infty$, d_p (如上题所定义的) 不可能在 $\mathcal{L}^p(X, \mathcal{S}, \mu)$ 上定义一个伪度量.
9. 设 $(X, \mathcal{S}, \mu)$ 是一个测度空间，f, g 是 X 上的正可测函数. 设 $0<t<r<m<\infty$.
   (a) 证明：Rogers 不等式 $\left(\int fg^r\mathrm{d}\mu\right)^{m-t}\leqslant\left(\int fg^t\mathrm{d}\mu\right)^{m-r}\left(\int fg^m\mathrm{d}\mu\right)^{r-t}$ 成立，如果不等式的右端积分有限. [提示：用 Rogers-Hölder 不等式 (定理 5.1.2).]
   (b) 反之，证明如何由 Rogers 不等式推得 Rogers-Hölder 不等式成立. [提示：令 $t=1$, $m=2$.]

(c)证明：对任意的 $m>1$，$\left(\int fg\mathrm{d}\mu\right)^m \leqslant \left(\int f\mathrm{d}\mu\right)^{m-1}\left(\int fg^m\mathrm{d}\mu\right)$．（"historical Hölder 不等式"）．

157

(d)证明：如何由(c)中的不等式推得 Rogers-Hölde 不等式．

5.2 L^p 空间的范数及完备性

对任意集合 X，其上的伪度量是 d，令 $x\sim y$ 当且仅当 $d(x,y)=0$，则容易证明 ~ 是一个等价关系．对每个 $x\in X$，令 $x^{\sim}:=\{y: y\sim x\}$．对 $x\in X$，令 $X^{\sim}$ 是所有等价集类 $x^{\sim}$ 的集合．令 $d(x^{\sim},y^{\sim}):=d(x,y)$，$\forall x,y\in X$．显然，$d$ 是明确定义的且是 $X^{\sim}$ 上的一个度量．如果 X 是具有半范数 $\|\cdot\|$ 的向量空间，则 $\{x\in X: \|x\|=0\}$ 是 X 的一个向量子空间 Z．对每个 $y\in X$，令 $y+Z:=\{y+z: z\in Z\}$，则显然 $y^{\sim}=y+Z$．空间 $X^{\sim}=\{x+Z: x\in X\}$ 通常称为商空间或因子空间 X/Z，自然也是一个向量空间，在其上定义范数 $\|\cdot\|$．对每个 $\mathcal{L}^p$ 空间，用这种方式定义的等价类的因子空间称为 L^p，即 $L^p(X,\mathcal{S},\mu):=\{f^{\sim}: f\in\mathcal{L}^p(X,\mathcal{S},\mu)\}$．注意到，如果 $g\geqslant 0$ 且 g 是可测的，则 $\int g\mathrm{d}\mu=0$ 当且仅当对于 μ，$g=0$ a. e.．所以 $h\sim f$ 当且仅当 $h=f$ a. e. 有时一个函数无定义和/或在一个零测集上无界，只要它几乎处处等于 $\mathcal{L}^p$ 中的一个实值函数，就也可以说它在 $\mathcal{L}^p$ 中．不论怎样，等价类的集族 L^p 都一样，因为每个等价类包含一个实值函数．等价类 $f^{\sim}$ 有时称为函数类．很多作者都将几乎处处相等的函数当作一个函数，并且不严格区分 $\mathcal{L}^p$ 和 L^p．

定义 X 是一个向量空间，$\|\cdot\|$ 是 X 上的范数，则 $(X,\|\cdot\|)$ 称为**线性赋范空间**(normed linear space)．**巴拿赫空间**(Banach space)是一个完备的线性赋范空间，其上的度量是由范数来定义的．对于复数域 $\mathbb{C}$ 上的向量空间，半范数的定义除了对复数 c 有 $\|cx\|=|c|\|x\|$ 成立外，其他的定义保持不变．

例：有限维空间 $\mathbb{R}^k$ 是巴拿赫空间，范数为通常的范数 $(x_1^2+x_2^2+\cdots+x_k^2)^{1/2}$ 或 $|x_1|+|x_2|+\cdots+|x_k|$．令 $C_b(X)$ 是拓扑空间 X 上的所有有界连续实值函数构成的空间，范数为 $\|f\|_\infty:=\sup\{|f|: x\in X\}$，则 $C_b(X)$ 是一个巴拿赫空间，在定理 2.4.9 中已经证明它的完备性．

5.2.1 定理（L^p 的完备性） 对任意测度空间 $(X,\mathcal{S},\mu)$ 且 $1\leqslant p\leqslant\infty$，$(L^p(X,\mathcal{S},\mu),\|\cdot\|_p)$ 是一个巴拿赫空间．

158

证明 正如在 Minkowski-Riesz 不等式（定理 5.1.5）后所定义的，$\|\cdot\|_p$ 是 $\mathcal{L}^p$ 上的半范数，容易得出它在 L^p 上定义一个范数．为了证明它是完备的，令 $\{f_n\}$ 是 $\mathcal{L}^p$ 上的柯西序列．如果 $p=\infty$，则对几乎所有的 x，有 $|f_m(x)-f_n(x)|\leqslant\|f_m-f_n\|_\infty$ 对所有的 m，n 成立（对可数多个对 (m,n) 取零测集的一个并）．对这样的 x，$f_n(x)$ 收敛到某个数（如 $f(x)$）．对 x 的其他值，令 $f(x):=0$，则由定理 4.2.2 知，f 是可测的．对几乎所有的 x 及所有的 m，当 m 足够大时，$|f(x)-f_m(x)|\leqslant\sup_{n\geqslant m}\|f_n-f_m\|_\infty\leqslant 1$，则 $\|f\|_\infty\leqslant\|f_m\|_\infty+1$，所以 $f\in\mathcal{L}^\infty$ 且当 $n\to\infty$ 时，$\|f-f_n\|_\infty\to 0$，正如所期望的．

□

现在令 $1\leqslant p<\infty$．在任意的度量空间中，一个柯西列和它的子列收敛到相同的极限，所以只要证明子列的收敛性．从而可以假设对所有的 n 和 $m>n$，有 $\|f_m-f_n\|_p<1/2^n$．令 $A(n):=\{x: |f_n(x)-f_{n+1}(x)|\geqslant 1/n^2\}$，则 $1_{A(n)}/n^2\leqslant|f_n-f_{n+1}|$，所以对所有的 n，有 $\mu(A(n))/n^{2p}\leqslant\int|f_n-f_{n+1}|^p\mathrm{d}\mu<2^{-np}$，并且 $\sum_n\mu(A(n))\leqslant\sum_n n^{2p}/2^{np}<\infty$．从而对 $B(n):=\bigcup_{m\geqslant n}A(m)$，$B(n)\downarrow$

并且 $n\to\infty$ 时，$\mu(B(n))\to 0$. 对任意 $x\notin\bigcap_{n=1}^{\infty}B(n)$，对几乎所有的 x，以及所有足够大的 n，$|f_n(x)-f_{n+1}(x)|\leqslant 1/n^2$，$n\to\infty$. 则对任意 $m>n$，$|f_m(x)-f_n(x)|\leqslant\sum_{j=n}^{\infty}1/j^2$. 因为 $\sum_{j=1}^{\infty}1/j^2$ 收敛，当 $n\to\infty$ 时，$\sum_{j=n}^{\infty}1/j^2\to 0$. 所以对这样的 x，$\{f_n(x)\}$ 是一个柯西列，极限是 $f(x)$. 对形成零测集的其他 x，令 $f(x)=0$. 则正如前面所定义的，f 是可测的．由法图引理(4.3.3)，$\int|f|^p\mathrm{d}\mu\leqslant\liminf_{n\to\infty}\int|f_n|^p\mathrm{d}\mu<\infty$（任意柯西列是有界的）. 同样，当 $n\to\infty$ 时，$\int|f-f_n|^p\mathrm{d}\mu\leqslant\liminf_{m\to\infty}\int|f_m-f_n|^p\mathrm{d}\mu\to 0$，所以 $\|f_n-f\|_p\to 0$.

例：对 $1\leqslant p<\infty$，存在序列 $\{f_n\}\subset\mathcal{L}^\infty([0,1],\mathcal{B},\lambda)$，并且在 $\mathcal{L}^p$ 中 $f_n\to 0$，但对任意的 x，$f_n(x)$ 不一定收敛到 0. 令 $f_1:=1_{[0,1]}$，$f_2:=1_{[0,1/2]}$，$f_3:=1_{[1/2,1]}$，$f_4:=1_{[0,1/3]}$，…，$f_k:=1_{[(j-1)/n,j/n]}$，其中 $k=j+n(n-1)/2$，$n=1,2,\cdots$，$j=1,\cdots,n$，则 $f_k\geqslant 0$，对 $1\leqslant p<\infty$，当 $k\to 0$ 时，$\int f_k(x)^p\mathrm{d}\lambda(x)=n^{-1}\to 0$，然而对所有的 x，有

$$\liminf_{k\to\infty}f_k(x)=0<1=\limsup_{k\to\infty}f_k(x).$$

在 L^p 的完备性证明中，这表明了为什么要得到一个逐点收敛的子列．

习题

1. 对 $0<p<1$，$\mathcal{L}^p(X,\mathcal{S},\mu)$ 是 X 上的所有可测实函数的集合且 $\int|f|^p\mathrm{d}\mu<\infty$，伪度量 $d_p(f,g)=\int|f-g|^p\mathrm{d}\mu$. 证明：$\mathcal{L}^p$ 关于 d_p 是完备的．

159

2. 设 $(X,\mathcal{S},\mu)$ 是一个测度空间，$(Y,\|\cdot\|)$ 是一个巴拿赫空间．证明：$\mathcal{L}^p(X,\mathcal{S},\mu)$（如 5.1 节习题 4 中所定义的）是完备的．
3. 对于 $\mathcal{L}^p$ 的完备性的另一种证明，找一个满足定理 5.2.1 证明中的子列后，令 f_n 是 $\mathcal{L}^p(X,\mathcal{S},\mu)$ 中的一个函数列，这里 $1\leqslant p\leqslant\infty$，使得 $\sum_{n\geqslant 1}\|f_n-f_{n+1}\|_p<\infty$. 令 $G_n:=\sum_{1\leqslant k\leqslant n}|f_{k+1}-f_k|$. 证明：
   (a) 序列 G_n 几乎处处收敛到 $\mathcal{L}^p$ 中的一个函数 G.
   (b) 在 $\mathcal{L}^p$ 中 G_n 收敛到 G.
   (c) 当 $\{G_n\}$ 收敛时，$\{f_n\}$ 也收敛.
   (d) 令 $f_n\to f$ a. e.，用控制收敛定理证明在 $\mathcal{L}^p$ 中，$f_n\to f$.
4. 对 $1\leqslant p<r<+\infty$，λ 为 $\mathbb{R}$ 上的勒贝格测度，令 $S:=\mathcal{L}^p(\mathbb{R},\lambda)\cap\mathcal{L}^r(\mathbb{R},\lambda)$ 且 $\|\cdot\|:=\|\cdot\|_p+\|\cdot\|_r$. 证明或举反例证明：$(S,\|\cdot\|)$ 是完备的．
5. 从实向量空间 V 映射到 $[0,\infty)$ 是一个函数 p 称为*正齐次的*（positive homogeneous），如果对每个 $v\in V$，$c\geqslant 0$，$p(cv)=cp(v)$，并称为*次可加的*当且仅当对 V 中的任意 u，v，有 $p(u+v)\leqslant p(u)+p(v)$.
   (a) 举一个 $\mathbb{R}$ 上的次可加正齐次函数的例子，此函数仅在 0 处为 0，但此函数不是半范数也不是范数.
   (b) 如果 p 是实向量空间 X 上的次可加正齐次函数，证明：$\|x\|:=p(x)+p(-x)$ 是 X 上的一个半函数.
6. 设 $(X_i,\|\cdot\|_i)$ $(i=1,2,\cdots)$ 是一个巴拿赫空间．直和 X 定义为所有序列 $x:=\{x_i\}_{i\geqslant 1}$ 的集合，使得对所有的 i，$x_i\in X_i$ 并且 $\|x\|:=\sum_i\|x_i\|_i<\infty$. 证明：$(X,\|\cdot\|)$ 是一个巴拿赫空间.
7. （连续性）. 对每个 $i=1,2,\cdots$，设 X_j 是二维空间 $\mathbb{R}^2$，其范数为通常范数 $\|y_j,z_j\|:=(y_j^2+z_j^2)^{1/2}$. 设 U_j，

V_j 是 X_j 的一维子空间，这里 $U_j=\{(c,0): c\in\mathbb{R}\}$，$V_j=\{(jc,c): c\in\mathbb{R}\}$. 设 U 是 U_j 的直和，V 和 V_j 的直和，X 是 X_j 的直和. 证明：U 和 V 是完备的，但 $U+V:=\{u+v: u\in U, v\in V\}$ 在 X 中不完备. [提示：考虑 $\{(0,j^{-2})\}_{j=1}^{\infty}$.]

5.3 希尔伯特空间

设 H 是数域 K 上的向量空间，这里 $K=\mathbb{R}$ 或 $\mathbb{C}$，$\mathbb{C}$ 是复数域. 对 $z=x+iy\in\mathbb{C}$（其中 x，y 是实数），$\bar{z}:=z^{-}:=x-iy$ 称为 z 的复共轭(complex conjugate，参见附录 B). H 的半内积(semi-inner product)是一个从 $H\times H$ 映射到 K 的函数$(\cdot,\cdot)$，使得

(i)对所有的 $c\in K$，f，$g\in H$，$(cf+g,h)=c(f,h)+(g,h)$.

160

(ii)对所有的 f，$g\in H$，$(f,g)=(g,f)^{-}$.（所以如果 $K=\mathbb{R}$，$(f,g)=(g,f)$.）

(iii)$(\cdot,\cdot)$是非负正定的，即对所有的 $f\in H$，$(f,f)\geqslant 0$.

对实数的情况，$H\times H$ 上的半内积满足(i)双线性，(ii)对称性，(iii)正定性. 对复数的情况，条件(ii)被称为"共轭对称". 那么由(i)和(ii)，$(h,cf+g)=\bar{c}(h,f)+(h,g)$. 所以$(h,f)$关于 h 是线性的，关于 f 是线性共轭的.

一个半内积称为内积(inner product)当且仅当$(f,f)=0$ 蕴涵着 $f=0$. 一个(半)内积空间是一个内积对$(V,(\cdot,\cdot))$，其中 V 是一个 $K=\mathbb{R}$ 或 $\mathbb{C}$ 上的向量空间，$(\cdot,\cdot)$是 V 上的(半)内积. 一个经典的内积例子就是向量空间的点积：对 z，$\omega\in K^n$，令 $(z,\omega):=\sum_{1\leqslant j\leqslant n} z_j\bar{\omega}_j$. 内积的一些主要例子在本章的引言中给出了，下面给出半内积的例子.

5.3.1 定理 对任意测度空间$(X,\mathcal{S},\mu)$，

$$(f,g):=\int f\bar{g}\,\mathrm{d}\mu$$

定义了实数空间或复数空间$\mathcal{L}^2(X,\mathcal{S},\mu)$上的一个半内积.

证明 首先考虑实数的情况. 那么(f,g)总是有定义的，由推论 5.1.4 知，(f,g)是有限的. 容易验证性质(i)、(ii)和(iii)成立. 对复数的情况，假设 $f=g+ih$ 和 $\gamma=\alpha+i\beta$ 属于复$\mathcal{L}^2$. 其中 g，h，α 和 β 都属于实$\mathcal{L}^2$. 那么通过应用四次实数的情况，(f,γ)是有定义的且有限的. 而且容易验证定义的半内积的性质. □

给定一个半内积$(\cdot,\cdot)$，令 $\|f\|:=(f,f)^{1/2}$. 正如此记号所表明的，下面将验证 $\|\cdot\|$ 是一个半范数.

5.3.2 命题 对所有的 $c\in K$，$f\in H$，$\|cf\|=|c|\|f\|$.

证明

$$\|cf\|=(cf,cf)^{1/2}=(c\bar{c}(f,f))^{1/2}=(|c|^2(f,f))^{1/2}=|c|\|f\|. \quad \square$$

推论 5.1.4 是形如 $\int fg\,\mathrm{d}\mu$ 的内积的一个柯西不等式，和在定理 5.3.1 中的一样. 这个不等式是很用的，而且对于半内积，下面的不等式成立.

161

5.3.3 Cauchy-Bunyakovsky-Schwarz 内积不等式 对任意半内积$(\cdot,\cdot)$和 f，$g\in H$，

$$|(f,g)|^2\leqslant(f,f)(g,g).$$

证明 由命题 5.3.2，f 可以乘以任意一个满足 $|c|=1$ 的复数，并且不改变不等式的方向. 所以假设(f,g)是实数，则对所有实数 x，$0\leqslant\|f+xg\|^2=(f,f)+2x(f,g)+x^2(g,g)$. 一个非

负的二项式 ax^2+bx+c，满足 $b^2-4ac\leqslant 0$，所以 $4(f,g)^2-4(f,f)(g,g)\leqslant 0$. □

现在证明半范数的性质.

5.3.4 定理 对任意(半)内积$(\cdot,\cdot)$和 $\|x\|:=(x,x)^{1/2}$，$\|\cdot\|$是一个(半)内积.

证明 用5.3.3可以得到，对任意的 $f,g\in H$，有

$$\|f+g\|^2=\|f\|^2+(f,g)+(g,f)+\|g\|^2\leqslant(\|f\|+\|g\|)^2.$$

由此和命题5.3.2得到所证结论. □

定义 $(H,(\cdot,\cdot))$是一个内积空间(inner product space)，当且仅当 H 是 $K=\mathbb{R}$ 或 $\mathbb{C}$ 上的一个向量空间并且$(\cdot,\cdot)$是 H 上的一个内积. 那么 H 称为**希尔伯特空间**(Hilbert space)，如果 H(对范数定义的度量 $d(x,y):=\|x-y\|=(x-y,x-y)^{1/2}$是完备的.

例：令$(X,\mathcal{S},\mu)$是一个测度空间. 则 $L^2(X,\mathcal{S},\mu)$是一个希尔伯特空间，内积为定理5.3.1中定义的$(f^{\sim},g^{\sim}):=(f,g):=\int f(x)\bar{g}(x)\mathrm{d}\mu(x)$. 由定理5.2.1知，它是完备的. 如果 X 是任意集合，其上的测度为计数测度，则$\mathcal{L}^p$和 L^p 完全一样，称为 $\ell^p(X)$. 特别地，如果 X 是正整数集，$\ell^p(X)$称为 ℓ^p. 那么 ℓ^2 是希尔伯特空间，集合是所有的“平方和”序列$\{z_k\}$，序列$\{z_k\}$满足 $\sum_k |z_k|^2<\infty$，并且 $(\{z_k\},\{w_k\})=\sum_k z_k\bar{w}_k$.

从定理5.3.3易知，内积(x,y)对固定的 y，关于 x 是连续的. (x,y)的联合连续性需要下面的证明.

5.3.5 定理 对任意的内积空间$(H,(\cdot,\cdot))$，内积是联合连续的，即从 $H\times H$ 映射到 K 的乘积拓扑是连续的.

162

证明 给定 $u,v\in H$，由不等式5.3.3和定理5.3.4知，当 $x\to u$，$y\to v$ 时，$|(x,y)-(u,v)|\leqslant|(x,y-v)|+|(x-u,v)|\leqslant\|x\|\|y-v\|+\|x-u\|\|v\|\leqslant(\|u\|+\|x-u\|)\|y-v\|+\|x-u\|\|v\|\to 0$，因为 $\|x-u\|\to 0$，$\|y-v\|\to 0$. □

我们说 H 的两个元素 f,g 是正交的(orthogonal)或垂直的(perpendicular)，记为 $f\perp g$，当且仅当 $(f,g)=0$. 例如，在 $L^2(X,\mathcal{S},\mu)$空间中，不相交集合的示性函数是正交的. 下面的两个事实扩展了平面几何的经典定理，(用当前的定义)这是很直接的.

5.3.6 巴比伦-毕达哥拉斯定理 如果 $f\perp g$，则 $\|f+g\|^2=\|f\|^2+\|g\|^2$.

5.3.7 平行四边形法则 对任意由半内积定义的半范数$\|\cdot\|$，及任意的 $f,g\in H$，

$$\|f+g\|^2+\|f-g\|^2=2\|f\|^2+2\|g\|^2.$$

对任意 $A\subset H$，令 $A^{\perp}:=\{y:(x,y)=0,\ \forall x\in A\}$. 所以在$\mathbb{R}^2$平面上，如果 A 是过原点的直线，则 $A^{\perp}$是垂线. 对于平面中的任意点 P，我们能作出一条 A 的垂线，它和 A 上的一个点 y 相交. 向量 $z=p-y\in A^{\perp}$，所以 $p=y+z$，p 可分解为两个正交向量的和. 现在这种分解可以扩展到内积空间，使得 A 是 F 的一个线性子空间. 注意，F 的完备性需要用几何构造来证明.

5.3.8 定理(正交分解) 对任意的内积空间$(H,(\cdot,\cdot))$，希尔伯特空间(即完备的线性空间)的子空间 $F\subset H$，且 $x\in H$，则 $y\in F$，$z\in F^{\perp}$有一个唯一的表示 $x=y+z$.

证明 令 $c:=\inf\{\|x-f\|:f\in F\}$. 取$\{f_n\}\in F$，满足 $\|x-f_n\|\downarrow c$. 那么对 f_m-x，f_n-x 应用平行四边形法则，对任意的 m,n，

$$\|f_m-f_n\|^2=2\|f_m-x\|^2+2\|f_n-x\|^2-4\|(f_m+f_n)/2-x\|^2.$$

因为$(f_m+f_n)/2\in F$，$\|(f_m+f_n)/2-x\|\geqslant c$，所以当 $m,n\to\infty$ 时，$\|f_m-f_n\|\to 0$. 假设 F 是完备

的，所以对某个 $y\in F$，当 $n\to\infty$ 时，$f_n\to y$. 令 $z:=x-y$，由连续性(定理 5.3.5)，$\|z\|=c$. 如果 $\exists f\in F$ 有 $(z,f)\neq 0$，令 $u:=(z,f)v$，$v\downarrow 0$，那么当 v 足够小时，

$$\|x-(y+uf)\|^2=\|z-uf\|^2=c^2-2v|(z,f)|^2+v^2|(z,f)|^2\|f\|^2<c^2,$$

163

而 $y+uf\in F$，和 c 的定义相矛盾. 所以 $z\in F^\perp$，如果还有 $x=g+h$，$g\in F$，$h\in F^\perp$，则 $y-g+z-h=0$. 且由定理 5.3.6，$0=\|y-g+z-h\|^2=\|y-g\|^2+\|z-h\|^2$，所以 $y=g$，$z=h$. □

注：对希尔伯特空间 H 中的每个 x，H 的闭线性子空间 F，令 $\pi_F(x):=y\in F$，这里 $x=y+z$ 是 x 的正交分解，则 π_F 是线性的，称为从 H 到 F 的正交投影.

$\mathcal{L}^2$ 上的具有勒贝格测度的很多函数是无界的，不是黎曼可积的. 所以黎曼可积的函数空间在 $\mathcal{L}^2$ 范数下是不完备的，并且正交分解不可行. 这就表明了勒贝格积分的优势.

习题

1. 在 $\mathbb{R}^2$ 中，内积为通常的内积 $((x,y),(u,v)):=xu+yv$. 令 $F:=\{(x,y): x+2y=0\}$. 问 $F^\perp$ 是什么形式的？
2. 在具有通常内积的 $\mathbb{R}^3$ 空间中，令 $F:=\{(x,y,z): x+2y+3z=0\}$. 令 $u:=(1,0,0)$. 求 u 的正交分解，使得 $u=v+w$，$v\in F$，$w\in F^\perp$.
3. 向量空间中的集合 C 称为凸的(convex)当且仅当对 C 中的所有 x，y 及 $0\leqslant t\leqslant 1$，有 $tx+(1-t)y\in C$，假设 C 是希尔伯特空间 H 中的一个闭凸集. 证明：对每个 $u\in H$，有唯一一个最近的点 $v\in C$，换句话说，对 C 中的任意 w，$w\neq v$，有 $\|u-w\|>\|u-v\|$. [提示：参见正交分解定理(5.3.8)的证明.]
4. 在具有通常度量的 $\mathbb{R}^2$ 中，令 C 是集合 C: $=\{(x,y): x^2+2x+y^2-2y\leqslant 3\}$. 设 $u:=(1,2)$. 在 C 中找最接近于点 u 的点，证明它是最近的.
5. $n\times n$ 的矩阵 $A:=\{A_{jk}\}_{1\leqslant j\leqslant n,1\leqslant k\leqslant n}$，并且所有的 $A_{jk}\in K$ 通过 $A(\{z_j\}_{1\leqslant j\leqslant n}):=\left\{\sum_{1\leqslant k\leqslant n}A_{jk}z_k\right\}_{1\leqslant j\leqslant n}$ 定义了一个从 K^n 映射到自身的一个线性变换. 那么 A 是非负定的，如果对所有的 $z\in K^n$，$(Az,z)\geqslant 0$. 证明：对 $K=\mathbb{C}$，如果 A 是非负定的，则对所有的 j，k，$A_{jk}=\bar{A}_{kj}$，但是如果 $K=\mathbb{R}$，这个性质不一定成立(A 不需要是对称的).
6. 证明：如果$(X,\|\cdot\|)$是一个实的赋范线性空间，其中范数满足平行四边形法则(如 5.3.7 中的)，对 X 中的所有 x，y，有 $\|x+y\|^2+\|x-y\|^2=2\|x\|^2+2\|y\|^2$，则存在一个内积$(\cdot,\cdot)$，范数为对所有的 $x\in X$，$\|x\|=(x,x)^{1/2}$. [提示：可以考虑 $\|x+y\|^2-\|x-y\|^2$.]

164

7. 在测度为勒贝格测度 λ 的$[0,1]$上，证明：对 $1\leqslant p\leqslant\infty$，在 $L^p([0,1],\lambda)$上，对某个内积$(\cdot,\cdot)$有范数 $\|f\|_p=(f,f)^{1/2}$ 当且仅当 $p=2$. [提示：用平行四边形法则 5.3.7.]
8. 在 $L^2([-1,1],\lambda)$中，令 F 是 f 的集合，f 满足对几乎所有的 x，有 $f(x)=f(-x)$. 证明：F 是闭线性子空间，$F^\perp$ 由所有的 g 组成，g 满足对几乎所有的 x，$g(-x)=-g(x)$. 令 $h(x)=(1+x)^4$，求 $f\in F$，$g\in F^\perp$，使得 $h=g+f$.
9. 令 X 是一个非零复向量空间. 求在复数 z，w 满足什么条件下，对 X 上的任意两个内积$(\cdot,\cdot)$和$((\cdot,\cdot))$，$A(u,v):=z(u,v)+w(u,v)$ 总是一个内积？
10. 令 X 是一个非零复值向量空间. 在复数 u，v 和 w 满足什么条件下，对 X 上的任意两个内积$(\cdot,\cdot)$和$((\cdot,\cdot))$，$B(f,g):=u(f,g)+v(g,f)+w(f,g)$ 总是一个内积？
11. 令 F 是使得 $\sup_n n|x_n|<\infty$ 成立的所有序列 $\{x_n\}$ 的集合. 证明或举反例证明：

 (a) F 是 ℓ^2 的一个线性子空间.

 (b) ℓ^2 中 F 的交集在 ℓ^2 中是闭的.

5.4 规范正交集和规范正交基

半内积空间 H 中的一个集合 $\{e_i\}_{i\in I}$ 称为规范正交的(orthonormal)当且仅当若 $i=j$，则 $(e_i, e_j)=1$，否则，$(e_i, e_j)=0$. 例如，$\mathbb{R}^n$ 中的单位向量 e_i 是规范正交基，这里 e_i 表示第 i 个坐标为 1，其余坐标为 0. 在 $\mathbb{R}^2$ 中，另一个规范正交基为 $e_1=2^{-1/2}(1, 1)$，$e_2=2^{-1/2}(-1, 1)$. 通过归纳把巴比伦-毕达哥拉斯定理(5.3.6)可以扩展到有限项，给出下面的事实.

5.4.1 定理 对任意有限集 $\{u_1, \cdots, u_n\}\subset H$，$u_i\perp u_j$，只要 $i\neq j$，就有 $\left\|\sum_{1\leqslant j\leqslant n}u_j\right\|^2=\sum_{1\leqslant j\leqslant n}\|u_j\|^2$. 因此，对任意有限规范正交集 $\{e_j\}_{1\leqslant j\leqslant n}$，$z_j\in K=\mathbb{R}$ 或 $\mathbb{C}$

165

$$\left\|\sum_{1\leqslant j\leqslant n}z_je_j\right\|^2=\sum_{1\leqslant j\leqslant n}|z_j|^2.$$

回想一个拓扑空间 $(X, \mathcal{T})$ 称为豪斯多夫空间，当且仅当对 X 中任意的 $x\neq y$，有不相交的开集 U 和 V，且 $x\in U$，$y\in V$. 在这样一个空间中，2.1 节中所定义的收敛网格有唯一的极限. 给定指标集 I 和一个从 I 映射到向量空间 X 的函数 $i\mapsto x_i$，X 上的拓扑是豪斯多夫拓扑 $\mathcal{T}$，和 $\sum_i x_i$ 是如下定义的. 令 $\mathcal{F}$ 是 I 的所有有限子集的集族，则 $\sum_{i\in I}x_i=x$ 当且仅当对 $\mathcal{T}$，集族 $\left\{\sum_{i\in F}x_i\right\}_{F\in\mathcal{F}}$ 收敛到 x. 这种收敛称为无条件收敛，因为它与 I 中元素的顺序无关. 如果 $I=\mathbb{N}$，则 $\sum_n x_n=x$ 是有定义的，意味着当 $m\to\infty$ 时，关于 $\mathcal{T}$ $\sum_{0\leqslant j\leqslant m}x_j\to x$. 则像 $\sum_n(-1)^n/n$ 这样的级数是收敛的但不是无条件收敛的.

在 $\mathbb{R}$ 中，$\sum_{\alpha\in I}x_\alpha=x\in\mathbb{R}$ 意味着对任意 $\varepsilon>0$，有一个有限集 $F\subset I$，使得对任意的有限集 G 且 $F\subset G\subset I$，$\left|x-\sum_{\alpha\in G}x_\alpha\right|<\varepsilon$. $\sum_{\alpha\in I}x_\alpha=+\infty$ 意味着对任意有限的 M，有一个有限的 $F\subset I$，使得对任意有限的 G，且 $F\subset G\subset I$，$\sum_{\alpha\in G}x_\alpha>M$.

根据引理 3.1.2(附录 D)，非负项级数可以以任意顺序相加. 下面的定理说明了同样的事实，但是对于不可数指标集.

5.4.2 引理 在 $\mathbb{R}$ 中，如果对所有的 $\alpha\in I$，有 $x_\alpha\geqslant 0$，则

$$\sum_{\alpha\in I}x_\alpha=S(\{x_\alpha\}):=\sup\left\{\sum_{\alpha\in F}x_\alpha:F\text{ 有限},F\subset I\right\}.$$

证明 首先假设 $S(\{x_\alpha\})<\infty$. 设 $\varepsilon>0$，有一个有限的 $F\subset I$ 并且 $\sum_{\alpha\in F}x_\alpha>S(\{x_\alpha\})-\varepsilon$. 则对任意有限的 $G\subset I$ 且 $F\subset G$，$S(\{x_\alpha\})-\varepsilon<\sum_{\alpha\in F}x_\alpha\leqslant\sum_{\alpha\in G}x_\alpha\leqslant S(\{x_\alpha\})$. 这推出 $\sum_{\alpha\in I}x_\alpha=S(\{x_\alpha\})$.

如果 $S(\{x_\alpha\})=+\infty$，则对任意 $M<\infty$，有一个有限的 $F\subset I$ 且 $\sum_{\alpha\in F}x_\alpha>M$，对任意有限的 $F\subset G$，用 G 替代 F 有同样的结论成立，所以由定义有 $\sum_{\alpha\in I}x_\alpha=+\infty$. □

假设指标集 I 是不可数的，从而有不可数多个不同于 0 的实数 x_α. 则对某个 n，要么有无限多个 α，使得 $x_\alpha>1/n$，要么有无限多个 α，使得 $x_\alpha<-1/n$，因为有限集的可数并是可数的. 另一种情况，$\sum_{\alpha\in I}x_\alpha$ 可能不收敛到任意有限的极限. 实际上，实数的和通常包含可数的指标集，如有限集 $\mathbb{N}$，或正整数集. 如果它包含不可数的指标集，则可以处理为除和中的可数项外其余为 0. 下面的两个事实将内积和规范正交基的坐标及规范正交基的和联系起来.

166

5.4.3 贝塞尔不等式 对任意的规范正交集$\{e_\alpha\}_{\alpha\in I}$及$x\in H$，有$\|x\|^2\geqslant\sum_{\alpha\in I}|(x,e_\alpha)|^2$.

证明 由引理5.4.2可以假设I是有限的．那么令$y:=\sum_{\alpha\in I}(x,e_\alpha)e_\alpha$. 现在由内积的线性性和定理5.4.1知，$(x,y)=\sum_{\alpha\in I}|(x,e_\alpha)|^2=(y,y)$. 因此，$x-y\perp y$. 所以由定理5.3.6，$\|x\|^2=\|x-y\|^2+\|y\|^2\geqslant\|y\|^2=(y,y)$. □

在$\mathbb{R}^n$中，两个向量$x=(x_1,\cdots,x_n)$和$y=(y_1,\cdots,y_n)$的内积通常定义为$(x,y)=x_1y_1+\cdots+x_ny_n$. 对于$\mathbb{R}^n$中通常的规范正交基$e_1,\cdots,e_n$，内积也可记为$\sum_{1\leqslant j\leqslant n}(x,e_j)(y,e_j)$. 后一种和的写法表明内积与$\mathbb{R}^n$中规范正交基的选择无关，可以扩展到如下一般的内积空间中的规范正交集.

5.4.4 定理(帕塞瓦尔–贝塞尔不等式) 如果$\{e_\alpha\}$是H中的规范正交集，$x\in H$，$y\in H$，$x=\sum_{\alpha\in I}x_\alpha e_\alpha$且$y=\sum_{\alpha\in I}y_\alpha e_\alpha$，则对所有的$\alpha$，$x_\alpha=(x,e_\alpha)$，$y_\alpha=(y,e_\alpha)$且$(x,y)=\sum_{\alpha\in I}x_\alpha\bar{y}_\alpha$.

证明 由内积的连续性(定理5.3.5)知，对每个$\beta\in I$，$(x,e_\beta)=\left(\sum_{\alpha\in I}x_\alpha e_\alpha,e_\beta\right)=x_\beta$，对$y$有同样的结果.

利用内积的线性性质，只要有限多个x_α和y_α不为0，就有定理中的最后一个式子成立．对任意有限的$F\subset I$，令$x_F:=\sum_{\alpha\in F}x_\alpha e_\alpha$，$y_F:=\sum_{\alpha\in F}y_\alpha e_\alpha$. 那么网格$x_F\to x$，$y_F\to y$. 再由内积的连续性(定理5.3.5)，$(x_F,y_F)\to(x,y)$，即推出了所要的结果. □

5.4.5 Riesz-Fischer 定理 对任意的希尔伯特空间H，任意的规范正交集$\{e_\alpha\}$和任意$x_\alpha\in K$，$\sum_{\alpha\in I}|x_\alpha|^2<\infty$当且仅当$\sum_{\alpha\in I}x_\alpha e_\alpha$在$H$中收敛到某个$x$.

证明 "充分性"可由不等式5.4.3和定理5.4.4得到．"必要性"：对每个$n=1,2,\cdots$选择一个有限集$F(n)\subset I$，使得$F(n)$随着n的增大而增大且$\sum_{\alpha\notin F(n)}|x_\alpha|^2<1/n^2$. 于是由定理5.4.1，$\{\sum_{\alpha\in F(n)}x_\alpha e_\alpha\}$是柯西序列，收敛到某个$x$，因为$H$是完备的．则所有的部分和网格收敛到同一个极限. □

在ℓ^2空间中，所有的序列$x=\{x_n\}_{n\geqslant1}$，满足$\sum_n|x_n|^2<\infty$，设$\{e_n\}_{n\geqslant1}$是标准基，即对于$j=n$，$(e_n)_j=1$，其余情况为0. 那么Riesz-Fischer定理蕴涵着$\sum_n x_ne_n$收敛到ℓ^2中的x. 换句话说，如果对$j\leqslant n$，$(y^{(n)})_j=x_j$；对$j>n$，$(y^{(n)})_j=0$，则在ℓ^2中$y^{(n)}\to x$.（对比着，考虑所有有界序列$\{x_j\}_{j\geqslant1}$组成的空间ℓ^∞，其范数为上确界范数$\|x\|_\infty=\sup_j|x_j|$．那么$y^{(n)}$关于范数$\|\cdot\|_\infty$不收敛到x，即对所有的j，如果$x_j=1$.） 167

设S是由K拓展的一个向量空间H的子集，S的线性生成空间定义为所有和$\sum_{x\in F}z_xx$的集合，这里F是S的任意有限子集且对每个$x\in F$，$z_x\in K$. 回顾S是线性无关的当且仅当S中没有元素y属于线性生成空间$S/\{y\}$. 给定一个线性无关序列，可以在同一个线性生成空间中找到一个规范正交集.

5.4.6 格拉姆–施密特规范正交化定理 令H是任意的内积空间，且$\{f_n\}$是H中任意的线性无关序列．那么在H中有一个规范正交序列$\{e_n\}$，使得对每个n，$\{f_1,\cdots,f_n\}$和$\{e_1,\cdots,e_n\}$有相同的线

性生成空间．

证明 令 $e_1 := f_1 / \| f_1 \|$．只证对 $n=1$ 的情况．给定 e_1，…，e_{n-1}，由递归，令

$$g_n := f_n - \sum_{1 \leq j \leq n-1} (f_n, e_j) e_j.$$

那么对所有的 $j < n$，$g_n \perp e_j$，但 g_n 不为 0，因为 f_n 不在 $\{f_1, \cdots, f_{n-1}\}$ 生成的线性空间中．由归纳假设，这与由 $\{e_1, \cdots, e_{n-1}\}$ 生成的线性空间相同．令 $e_n := f_n / \| f_n \|$．于是所要证的性质成立． □

例如，在 $\mathbb{R}^3$ 中，令 $f_1 = (1, 1, 1)$，$f_2 = (1, 1, 2)$ 和 $f_3 = (1, 2, 3)$，则 $e_1 = 3^{-1/2}(1, 1, 1)$，$e_2 = 6^{-1/2}(-1, -1, 2)$ 和 $e_3 = 2^{-1/2}(-1, 1, 0)$．在这种情况下，f_1，f_2 只有一个一维正交子空间，所以在证明中不用计算就可以求得 e_3.

在有限维向量空间 S 中，一组基就是一个集合 $\{e_j\}_{1 \leq j \leq n}$，使得对某个数 s_j，S 中的每一个元素 s 仅可以唯一地表示为和 $\sum_{1 \leq j \leq n} s_j e_j$. 于是从 S 映射到它自身的线性变换 A 可表示为矩阵 $\{A_{jk}\}$ 且对每个 j，$A(e_j) = \sum_{1 \leq k \leq n} A_{jk} e_k$. 一般来说，有限维空间中的基是相当有用的．所以很自然地试图把这种基的思想扩展到无限维的空间中．但是在无限维空间中很难保持有限维空间中基的所有好的性质．在任意向量空间 S 中，一组哈梅尔(Hamel)基是一个使得每个 $s \in S$ 仅可以唯一地写为 $\sum_{\alpha \in I} s_\alpha e_\alpha$ 的集合 $\{e_\alpha\}_{\alpha \in I}$，其中仅有有限多的数 s_α 不为 0. 所以“哈梅尔基”是一个代数概念，并和 S 上的任意拓扑无关．对有限维空间，一组哈梅尔基仅是一组基，但是在无限维巴拿赫空间中(如希尔伯特空间)没有显然的哈梅尔基，虽然用选择性公理可以证明它是存在的．在分析中，通常要处理无限和的收敛，如泰勒级数或傅里叶级数，所以哈梅尔基通常不适用．在巴拿赫空间 $(S, \| \cdot \|)$ 中，无条件基(unconditional basis)是使得对每个 $s \in S$，存在唯一的数集 $\{s_\alpha\}_{\alpha \in I}$ 使得 $\sum_{\alpha \in I} s_\alpha e_\alpha = s$(对 $\| \cdot \|$ 收敛)的集族 $\{e_\alpha\}_{\alpha \in I}$. 在一个可分的巴拿赫空间 S 中，一组绍德尔(Schauder)基一个使得对每个 $s \in S$，有唯一的数列 s_n 使得 $\left\| s - \sum_{1 \leq j \leq n} s_j f_j \right\| \to 0 (n \to \infty)$ 的序列 $\{f_n\}_{n \geq 1}$. 在分析中最有用的可能是在可分的巴拿赫空间中找到绍德尔基，但是绍德尔基可能不是无条件基，一般来说，很难找到无条件基．一般的巴拿赫空间上的希尔伯特空间的若干优势之一就是希尔伯特空间有具有很好性质的基，下面给出它的定义.

168

定义 给定一个内积空间 $(H, (\cdot,\cdot))$，H 的一组规范正交基是一个规范正交集 $\{e_\alpha\}_{\alpha \in I}$，使得对每个 $x \in H$，$\sum_{\alpha \in I} (x, e_\alpha) e_\alpha$.

5.4.7 定理 每个希尔伯特空间有一组规范正交基．

证明 如果正交集的集族通过包含关系是线性序的，则它们的并也是一组规范正交集．因此，由佐恩引理(定理 1.5.1)，设 $\{e_\alpha\}_{\alpha \in I}$ 是最大的规范正交集．对任意的 $x \in H$，令 $y := \sum_{\alpha \in I} (x, e_\alpha) e_\alpha$，这里由贝塞尔不等式(5.4.3)和 Riesz-Fischer 定理(5.4.5)知，和是收敛的．如果 $y = x$，则我们已证明．否则，对所有的 α，有 $x - y \perp e_\alpha$，所以可以加入一个新的元素 $(x-y)/\|x-y\|$，这和规范正交集的最大性矛盾． □

例如，在希尔伯特空间 ℓ^2 中，“标准基” $\{e_n\}_{n \geq 1}$ 是规范正交的且很容易看出它是一组基．

给定两个度量空间 (S, d) 和 (T, e)，回顾等距同构就是一个从 S 映射到 T 的函数 f，使得对所有的 x，$y \in S$，$e(f(x), f(y)) = d(x, y)$. 如果 f 是映上 T 的，则 S 和 T 就是等距同构的．例如，在

具有通常度量的$\mathbb{R}^2$中，变换 $x \mapsto x+v$、围绕任意中心的旋转和任意直线上的投影都是等距同构的．

5.4.8 定理 每个希尔伯特空间 H 和一个集合 X 形成的空间 $\ell^2(X)$ 是等距同构的．

169

证明 令 $\{e_\alpha\}_{\alpha\in X}$ 是 H 的一组规范正交基．则由贝塞尔不等式(5.4.3)，$x \mapsto \{(x, e_\alpha)\}_{\alpha\in X}$ 使得 H 映射到 $\ell^2(X)$ 上．由帕塞瓦尔–贝塞尔不等式(5.4.4)知，这个函数保持内积．由 Riesz-Fischer 定理(5.4.5)知，此函数映上到 $\ell^2(X)$． □

定理5.4.8论述了每个希尔伯特空间都可以表示为计算测度的 L^2 空间，给出了一种相反事实就是每个 L^2 空间都是希尔伯特空间．下面的定理描述规范正交集之间基的特征．

5.4.9 定理 对任意的内积空间 $(H, (\cdot,\cdot))$，一个规范正交集 $\{e_\alpha\}_{\alpha\in I}$ 是 H 的一组规范正交基，当且仅当它的线性生成子空间在 H 中稠密．

证明 如果规范正交集是一组正交基，则显然 S 是稠密的．反之，假设 S 是稠密的．对每个 $x\in H$ 和有限集 $J\subset I$，令 $x_J := \sum_{\alpha\in J}(x, e_\alpha)e_\alpha$．那么对每个 $\alpha\in J$，$x - x_J \perp e_\alpha$．对任意 $c_\alpha \in K$，$z := \sum_{\alpha\in J} c_\alpha e_\alpha$，有 $x - x_J \perp x_J - z$．因此，由巴比伦–毕达哥拉斯定理(5.3.6)，

$$\| x - z \|^2 = \| x - x_J \|^2 + \| x_J - z \|^2 \geqslant \| x - x_J \|^2.$$

对任意 $\varepsilon > 0$，存在一个 J 和一个使得 $\| x - z\| < \varepsilon$ 的 z，于是 $\| x - x_J \| < \varepsilon$．如果 M 是 I 的另一个有限的子集且 $J\subset M$，则

$$x = x_J + (x_M - x_J) + (x - x_M),$$

其中 $x_M - x_J \perp x - x_M$．那么再由定理5.3.6，

$$\| x - x_J \|^2 = \| x_M - x_J \|^2 + \| x - x_M \|^2 \geqslant \| x - x_M \|^2.$$

因此，随着有限集 J 的增加，$\| x - x_J \|$ 收敛到0．换句话说，网格 $\{x_J\}$ 收敛到 x，且 $\{e_\alpha\}_{\alpha\in I}$ 是一组规范正交基． □

5.4.10 推论 设 $(H, (\cdot,\cdot))$ 是任意具有可数集的内积空间，且这个可数集有一个稠密的线性生成空间，则 H 有一组规范正交基．

证明 由格拉姆–施密特规范正交化定理(5.4.6)，有一个可数的规范正交集，这个可数的规范正交集有一个稠密的线性生成空间，由定理5.4.9知，它是一组基． □

170

注： 有稠密线性生成空间的可数集实际上和(有一个稠密集的)可分集是等价的，正如在线性生成空间中我们可以取有理系数且仍有一个稠密集．

对于勒贝格积分，L^2 空间已是完备的，但是如果对于连续函数，内积定义为 $(f, g) = \int_0^1 f(x)g(x)\,\mathrm{d}x$ (黎曼积分)，则需要得到一个希尔伯特空间才可证明完备性．

当度量空间完备时，度量 d 就可以扩张为完备的．对一个线性赋范空间，d 由范数 $\|\cdot\|$ 定义，当 $\|x\| = d(0, x)$ 时，度量 d 可以扩张为完备的．如果范数是由内积定义的，通过 $(x, y) = (\|x+y\|^2 - \|x-y\|^2)/4$ 在实数中扩充内积，在复数的情形有更复杂的公式．这里用一种不同的方法把内积扩张到完备的情形，以给出一个希尔伯特空间．

5.4.11 定理 设 $(J, (\cdot,\cdot))$ 是任意的内积空间．设 S 是 J 的完备空间(对通常的度量，范数是由内积定义的)．那么向量空间结构和内积都可扩张到 S 且是一个希尔伯特空间．

证明 设 H 是从 J 映射到 K 的所有连续共轭线性函数的集合．对 $f\in H$，设 $\|f\| := \sup\{\,|f(x)|\,:$

$x\in J$, $\|x\|\leqslant 1\}$，则 H 自然是 K 上的一个向量空间，$\|\cdot\|$ 是 H 上的一个范数．对 $j\in J$，设 $g(j):=(j,\ \cdot)$．那么由 Cauchy-Bunyakovsky-Schwarz 内积不等式(5.3.3)，g 是从 J 到 H 的映射且对所有 $j\in J$，$\|g(j)\|\leqslant\|j\|$．换句话说，对 J 中每个 $j\neq 0$，令 $x=j/\|j\|$，证明了 $\|g(j)\|\geqslant\|j\|$，所以 $\|g(j)\|=\|j\|$，且 g 是由 J 映射到 H 的等距同构．易见 H 是完备的．因此，g 的值域的闭包是 J 的完备化 S(这个闭包实际上是 H 的所有闭包，稍后将证明，这里用不到此结果)．内积的联合连续性和在它的证明中给出的强估计(定理 5.3.5)表明，通过连续性可以把内积从 J 扩张到 H 上(详细证明见习题 8)． □

例(傅里叶级数)：从历史上来看，某些三角函数集合可能是很好的，即使现在，它们仍是无限规范正交集的最重要的例子．令 $X=[-\pi,\ \pi)$，其上的测度为 $\mu:=\lambda/(2\pi)$，这里 λ 是勒贝格测度，则通过 $z=e^{i\theta}$，$-\pi<\theta<\pi$，X 可以等同于$\mathbb{C}$ 中的单位圆周$\{z:\ |z|=1\}$．对所有的整数 $n\in\mathbb{Z}$，X 上的函数 $f_n(\theta):=e^{in\theta}$是规范正交的．(这将在后面的命题 7.4.2 中证明，或用习题 6 中的另一种方法证明它们是 $L^2(X,\ \mu)$的一组基)．因此，对所有的 $n=1,\ 2\cdots$，常数函数 1、函数 $\theta\mapsto 2^{1/2}\sin(n\theta)$和 $\theta\mapsto 2^{1/2}\cos(n\theta)$形成了实 $L^2([-\pi,\ \pi),\ \mu)$的一组规范正交基(“实傅里叶级数”)．

171

习题

1. 设 $H=L^2([0,\ 1],\ \lambda)$，其中 λ 是$[0,\ 1]$上的勒贝格测度．设 $f_n(x):=x^{4+n}$，$n=1,\ 2,\ 3$. 求由格拉姆–施密特过程(5.4.6)给出的相应的规范正交基 e_1，e_2，e_3.
2. 设$(H,\ (\cdot,\cdot))$是内积空间，F 是一个完备的线性子空间．设$\{e_\alpha\}$是 F 的一组规范正交基．证明：$f(x):=\sum_\alpha(x,\ e_\alpha)e_\alpha$ 给出了从 x 到 F 的正交投影．
3. 如果$(H,\ (\cdot,\cdot))$是一个内积空间，$\{x_n\}$是一个序列且对所有的 $m\neq n$，$x_m\perp x_n$，证明：对 $\sum_{n\geqslant 1}x_n$，无条件收敛和一般的收敛是等价的(当 $n\to\infty$ 时序列 $\sum_{1\leqslant j\leqslant n}x_j$ 的收敛)．
4. 证明：$[0,\ 1]$上的连续实函数形成的空间 $C([0,\ 1])$在实 $L^2([0,\ 1],\ \lambda)$中是实稠密的．[提示：如果对所有的 $g\in C([0,\ 1])$，$(f,\ g)=0$，证明对每个可测集 A，有$(f,\ 1_A)=0$，开始证明时用区间 A.]
5. 条件同习题 4，证明：实 $L^2([0,\ 1],\ \lambda)$有一组规范正交基$\{P_n\}_{n\in\mathbb{N}}$，其中 P_n 是一个 n 阶的多项式(“勒让德(Legendre)多项式”)．[提示：用魏尔斯特拉斯逼近定理(2.4.12)和格拉姆–施密特过程．]
6. 条件同习题 4，证明：在定理 5.4.11 下面的例子中$\{\theta\mapsto e^{in\theta}\}_{n\in\mathbb{Z}}$是复空间 $L^2([0,\ 2\pi),\ \lambda/2\pi)$的一组规范正交基．[提示：用复斯通–魏尔斯特拉斯定理 2.4.13.]
7. 条件同习题 6，证明：函数 $\theta\mapsto\sqrt{2}\sin(n\theta)(n=1,\ 2,\ \cdots)$形成了实 $L^2([0,\ \pi),\ \lambda/\pi)$的一组规范正交基．[提示：考虑$[-\pi,\ \pi)$上的奇函数和偶函数.]
8. 详细证明：如何把内积空间上的内积扩张为它的完备化(定理 5.4.11 的证明)．
9. 设 f_n 是希尔伯特空间 H 上的规范正交列．当 α 取何值时，级数 $\sum_n f_n/(1+n^2)^\alpha$ 在 H 中收敛？
10. (Rademacher 函数)．对 $x\in[0,\ 1)$，当 $2k\leqslant 2^n x\leqslant 2k+1$ 时，令 $f_n(x):=1$，当 $2k+1\leqslant 2^n x<2k+2$ 时，令 $f_n(x):=-1$，对任意的整数 k，
    (a)证明：f_n 在 $H:=L^2([0,\ 1),\ \lambda)$中是规范正交的.
    (b)证明：f_n 不是 H 的基.

172

11. 对每个 $n=0,\ 1,\ \cdots,\ k=0,\ 1,\ \cdots,\ K(n)$，当 $2k\leqslant 2^n x\leqslant 2k+1$ 时，令 $f_{nk}(x):=a_{nk}$且当 $2k+1\leqslant 2^n x\leqslant 2k+2$ 时，令 $f_{nk}(x):=b_{nk}$．求 $K(n)$，a_{nk}及 b_{nk}的值；使得所有函数 f_{nk}的集合是 $L^2([0,\ 1),\ \lambda)$的一组规范正交基．如何唯一地确定 a_{nk}及 b_{nk}？

12. 设$(X, \mathcal{S}, \mu)$，$(Y, \mathcal{T}, \nu)$是两个测度空间．设$\{f_n\}$是$L^2(X, \mathcal{S}, \mu)$的规范正交基，$\{g_n\}$是$L^2(Y, \mathcal{T}, \nu)$的规范正交基．证明：所有函数$h_{mn}(x, y):=f_m(x)g_n(y)$的集合是$L^2(X\times Y, \mathcal{S}\otimes\mathcal{T}, \mu\times\nu)$的一组规范正交基．
13. 设H是一个具有规范正交基$\{e_n\}$的希尔伯特空间，令$h\in H$，$h=\sum_n h_n e_n$．证明：对$0\leqslant r\leqslant 1$，$\sum_n h_n r^n e_n$在H中收敛到一个$h^{(r)}$，其中当$r\uparrow 1$时，$h^{(r)}$收敛到h.
14. 在$\mathbb{C}$中的单位圆周$\{e^{i\theta}: -\pi<\theta\leqslant\pi\}$上，测度为$d\mu(\theta)=d\theta/(2\pi)$，由上面的习题6知，函数$\theta\mapsto e^{in\theta}(n\in\mathbb{Z})$．形成了一组规范正交基．在$H:=L^2(\mu)$中，令$H^2$是由函数$e^{in\theta}(n\geqslant 0)$生成的闭线性子空间．令$f(\theta):=\theta$，$-\pi<\theta\leqslant\pi$. 设$h$是$f$到$H^2$的正交投影．证明：$h$是无界的．[提示：用分部积分证明$f$有傅里叶级数$i\sum_{n\geqslant 1}(-1)^n(e^{in\theta}-e^{-in\theta})/n$.（对于$h$，此级数在$\theta=\pi$发散．）为了证明$h\notin\mathcal{L}^\infty(\mu)$，令$1+e^{i\theta}=\rho e^{i\varphi}$，其中$|\phi|\leqslant\pi/2$，$\varphi$是实数，且$\rho\geqslant 0$. 用习题13的方法证明对几乎所有的$\theta$，$h(\theta)=-i\cdot\log(1+e^{i\theta}):=-i\cdot\log\rho+\varphi$，证明如下．对$|z|<1$，$\log(1+z)$的泰勒级数（附录B）可写成$\log(1+re^{i\theta})(0\leqslant r<1)$的傅里叶级数．当$r\uparrow 1$时，证明在区间$|\theta|\leqslant\pi-\varepsilon(\varepsilon>0)$上一致收敛到$\log(1+e^{i\theta})$.（因为$H^2$是一个很自然的子空间，它上面的投影不保持有界性这一事实表明需要它是勒贝格可积的．）]

5.5 希尔伯特空间上的线性型、L^p 空间的包含关系及这两个度量之间的关系

令H是$K=\mathbb{R}$或$\mathbb{C}$上的一个希尔伯特空间．

从有限的欧几里得空间$\mathbb{R}^k$映射到$\mathbb{R}$的一个线性函数f总是连续的，可记为$f(x)=f(\{x_j\}_{1\leqslant j\leqslant k})=\sum_{1\leqslant j\leqslant k}x_j h_j=(x, h)$（$x$和$h$的内积）．这里$h_j=f(e_j)$，其中$e_j$是单位向量，单位向量就是第$j$个位置是1，其余是0的向量．连续的线性型的表示和内积一样可扩张到一般的希尔伯特空间.

173

5.5.1 定理（Riesz-Fréchet） *由希尔伯特空间H映射到K的一个函数f是线性连续的，当且仅当对某个$h\in H$，$f(x)=(x, h)$，$\forall x\in H$. 如果h存在，则h是唯一的．*

证明 “充分性”由定理5.3.5（连续性）和内积的定义（线性性）可以证明．为了证明“必要性”，设f是从H映射到K的连续线性函数．如果$f=0$，令$h=0$. 否则，令$F:=\{x\in H: f(x)=0\}$．因为f是连续的，F是闭的，因此F是完备的．选择任意$v\notin F$. 由正交分解（定理5.3.8），$v=y+z$，其中$y\in F$，$z\in F^\perp$，$z\neq 0$. 于是$f(z)\neq 0$. 设$u=z/f(z)$，则$f(u)=1$且$u\in F^\perp$. 对任意的$x\in H$，$x-f(x)u\in F$，所以$(x, u)-f(x)(u, u)=0$. 令$h:=u/(u, u)$，则对所有的$x\in H$，有$f(x)=(x, h)$.

如果对所有的$x\in H$，$(x, h)=(x, g)$，则$(x, h-g)=0$. 设置$x=h-g$，得$h=g$，则唯一性成立．

□

在一个有限测度空间上，对函数f，$|f|^p$的可积性受到p值的限制，p值越大对$|f|^p$的限制越大．在本节后面的定理中将用到这个结论．

5.5.2 定理 *设$(X, \mathcal{S}, \mu)$是任意的有限可测空间（$\mu(X)$有限）．那么对$1\leqslant r<s\leqslant\infty$，$\mathcal{L}^s(X, \mathcal{S}, \mu)\subset\mathcal{L}^r(X, \mathcal{S}, \mu)$，且从$L^s$映射到$\mathcal{L}^r$的恒等函数是连续的．*

证明 令$f\in\mathcal{L}^s$. 如果$s=\infty$，则$|f|\leqslant\|f\|_\infty$ a.e.，所以$\int|f|^r d\mu\leqslant\|f\|_\infty^r\mu(X)$，且$\|f\|_r\leqslant\|f\|_\infty\mu(X)^{1/r}$. 于是对$f, g\in\mathcal{L}^\infty$，$\|f-g\|_r\leqslant\|f-g\|_\infty\mu(X)^{1/r}$，这给出了连续性．如果$s<\infty$，则由Rogers-Hölder不等式（5.1.2）和$p:=s/r$，$q=p/(p-1)$得，$\int|f|^r d\mu=\int|f|^r\cdot 1d\mu\leqslant\left(\int|f|^{r(s/r)}d\mu\right)^{r/s}\mu(X)^{1/q}<\infty$. 因此，对所有的$f\in\mathcal{L}^s$，$\|f\|_r\leqslant\|f\|_s\mu(X)^{1/qr}$. 这就证明了连续性．

□

注： 如果$\mu(X)=1$，μ是概率测度，则对任意的可测函数f，随着p的增加，$\|f\|_p$关于p是非递减的．

现在令ν，μ是同一个可测空间$(X,\mathcal{S})$上的两个测度．称ν关于μ是绝对连续的(absolutely continuous)，记为$\nu<\mu$，当且仅当若$\mu(A)=0$，则$\nu(A)=0$. 称ν，μ都是奇异的(singular)，或$\nu\perp\mu$当且仅当有一个可测集A，满足$\mu(A)=\nu(X/A)=0$. 回顾对任意的测度μ，可测集A和任意可积或非负可测函数f，有$\int_A f\mathrm{d}\mu:=\int f1_A\mathrm{d}\mu$. 如果$f$是非负可测的，则$\nu(A):=\int_A f\mathrm{d}\mu$定义了一个关于$\mu$的绝对连续测度．如果$\mu$是$\sigma$-有限的，$\nu$是有限的，Radon-Nikodym定理(5.5.4)将证明这样的f的存在性对绝对连续是很有必要的．这是测度论中最重要的一个结论．下面两个定理将在一起证明．

174

5.5.3 定理(勒贝格分解)　设$(X,\mathcal{S})$是一个可测空间，μ，ν是其上的两个σ-有限测度．那么存在唯一的度量ν_{ac}，ν_s，使得$\nu=\nu_{ac}+\nu_s$，$\nu_{ac}<\mu$，$\nu_s\perp\mu$.

例如，设μ是$[0,2]$上的勒贝格测度，ν是$[1,3]$上的勒贝格测度，则ν_{ac}是$[1,2]$上的勒贝格测度，ν_s是$[2,3]$上的勒贝格测度．

5.5.4 定理(Radon-Nikodym)　在可测空间$(X,\mathcal{S})$上，设μ是σ-有限测度．设ν是关于μ绝对连续的有限测度．那么对某个$h\in\mathcal{L}^1(X,\mathcal{S},\mu)$，

$$\nu(E)=\int_E h\mathrm{d}\mu,\quad\forall E\in\mathcal{S}.$$

任意的这样两个h关于μ是几乎处处相等的．

定理5.5.3和定理5.5.4的证明　取一列不相交的可测集A_k，使得对所有的k，$X=\bigcup_k A_k$，$\mu(A_k)<\infty$. 同样，取关于ν的一列有限测度的不相交的可测集序列B_m，B_m的并是X. 那么取所有的$A_k\cap B_m$的交，就有一个不相交的可测的序列E_n，E_n的并是X，使得对所有的n，$\mu(E_n)<\infty$，$\nu(E_n)<\infty$.

对$\mathcal{S}$上的任意测度ρ，设对每个$C\in\mathcal{S}$，$\rho_n(C):=\rho(C\cap E_n)$. 那么$\rho(C)=\sum_n\rho_n(C)$，$\rho(C)=0$当且仅当对所有的$n$，$\rho_n(C)=0$. 因此，$\rho<\mu$当且仅当对所有的$n$，$\rho_n<\mu$，当且仅当对所有的$n$，$\rho_n<\mu_n$. $\rho\perp\mu$当且仅当对所有的n，$\rho_n\perp\mu$，当且仅当对所有的n，$\rho_n\perp\mu_n$.

对$\mathcal{S}$上的任意测度α_n且对所有的n，$\alpha_n(X/E_n)$，对每个$C\in\mathcal{S}$，$\alpha(C):=\sum_n\alpha_n(C)$定义了一个测度$\alpha$，$\alpha<\mu$当且仅当对所有的$n$，$\alpha_n<\mu_n$，且$\alpha\perp\mu$当且仅当对所有的$n$，$\alpha_n\perp\mu_n$. 因此，在勒贝格分解定理的证明中，$\mu$，$\nu$都可假设是有限的．对Radon-Nikodym定理(5.5.4)，如果在每个E_n上得到了证明，有某个函数$h_n\geqslant0$，设对每个$x\in E_n$，$h(x):=h_n(x)$. 那么h是可测的(引理4.2.4)且由定理5.5.4中的假设，

175

$$\int h\mathrm{d}\mu=\sum_{n\geqslant1}\int h_n\mathrm{d}\mu=\sum_{n\geqslant1}\nu(E_n)=\nu(X)<\infty,$$

所以$h\in\mathcal{L}^1(\mu)$. 因此，我们在定理5.5.4的证明中也假设μ(和ν一样)是有限的．

现在形成了希尔伯特空间$H=L^2(X,\mathcal{S},\mu+\nu)$. 由定理5.5.2知，$L^2\subset L^1$，且从$L^2$映射到$L^1$的恒等函数是连续的．并且，$\nu\leqslant\mu+\nu$. 因此，从$H$映射到$\mathbb{R}$的线性函数$f\mapsto\int f\mathrm{d}\mu$是连续的，则由定理5.5.1知，存在$g\in\mathcal{L}^2(X,\mathcal{S},\mu+\nu)$，使得对所有的$f\in\mathcal{L}^2(\mu+\nu)$，有$\int f\mathrm{d}\nu=\int fg\mathrm{d}(\mu+\nu)$. 因

此，对所有的 $f\in\mathcal{L}^2(\mu+\nu)$，

$$\int f(1-g)d(\mu+\nu) = \int f\mathrm{d}\mu \quad \forall f\in\mathcal{L}^2(\mu+\nu). \tag{*}$$

现在对 $\mu+\nu$，$g\geqslant 0$ a. e.，否则，取 f 为集合 $\{x: g(x)<0\}$ 的示性函数得到矛盾．同样，$(*)$ 式意味着 $g\leqslant 1$ a. e. $(\mu+\nu)$．因此，假设对所有的 x，$0\leqslant g(x)\leqslant 1$．于是由单调收敛定理，对所有可测的 $f\geqslant 0$，$(*)$ 式成立．

令 $A:=\{x: g(x)=1\}$，$B:=X\setminus A$．那么在 $(*)$ 式中令 $f=1_A$，使得 $\mu(A)=0$．对所有的 $E\in\mathcal{S}$，令

$$\nu_s(E):=\nu(E\cap A),\quad \nu_{ac}(E):=\nu(E\cap B).$$

因此 ν_s 和 ν_{ac} 是测度，$\nu=\nu_s+\nu_{ac}$，且 $\nu_s\perp\mu$．如果 $\mu(E)=0$，$E\subset B$，则由 $(*)$ 式及在 E 上 $1-g>0$，有 $\int_E(1-g)\mathrm{d}(\mu+\nu)=0$，所以 $(\mu+\nu)(E)=0$，$\nu(E)=\nu_{ac}(E)=0$．因此，$\nu_{ac}<\mu$．所以就证明了勒贝格分解的存在性．

为了证明唯一性，仍然假设两个测度是有限的．假设 $\nu=\rho+\sigma$，$\rho<\mu$，$\sigma\perp\mu$，则 $\rho(A)=0$，因为 $\mu(A)=0$．因此对所有的 $E\in S$，

$$\nu_s(E)=\nu(E\cap A)=\sigma(E\cap A)\leqslant\sigma(E).$$

所以 $\nu_s\leqslant\sigma$，$\rho\leqslant\nu_{ac}$．那么 $\sigma-\nu_s=\nu_{ac}-\rho$ 是一个测度，且都关于 μ 是绝对连续和奇异的，所以为 0，$\rho=\nu_{ac}$，$\sigma=\nu_s$．这就证明了勒贝格分解的唯一性，完成了定理 5.5.3 的证明．

现在来证明定理 5.5.4，由 $\nu=\nu_{ac}$，所以 $\nu_s=0$．令在 B 上 $h:=g/(1-g)$，在 A 上 $h:=0$．对任意的 $E\in\mathcal{S}$，在 $(*)$ 式中令 $f=h1_E$，则

$$\int_E h\mathrm{d}\mu=\int_{B\cap E}g\mathrm{d}(\mu+\nu)=\nu(B\cap E)=\nu(E).$$

从而在 Radon-Nikodym 定理中证明了 h 的存在性．

假设 j 是另一个使得对所有的 $E\in S$，$\int_E j\mathrm{d}\mu=\nu(E)$ 的函数，所以 $\int_E j-h\mathrm{d}\mu=0$．

令 $E_1:=\{x: j(x)>h(x)\}$，$E_2:=\{x: j(x)<h(x)\}$，在这些集合上对 $j-h$ 求积分可得 $\mu(E_1)=\mu(E_2)=0$．更详细地，可以考虑使得 $j-h>1/n$ 的集合，证明每个测度为 0，因此，它们的并的测度也为 0，同样地考虑 $j-h<-1/n$ 的集合．所以 $j=h$ a. e. (μ)，这就完成了定理 5.5.4 的证明． □

Radon-Nikodym 定理中的函数 h 称为 ν 关于 μ 的 Radon-Nikodym 导数（derivative）或密度（density），记为 $h=\mathrm{d}\nu/\mathrm{d}\mu$．习题中会证明这种记号的一些正确的结论． 176

习题

1. 证明：如果 ν 是一个有限的测度，并且关于 σ-有限测度 μ 是绝对连续的，f 是任意的关于 ν 可积的函数，则
$$\int f\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu=\int f\mathrm{d}\nu.$$
2. 如果 β，ν，μ 是有限测度，使得 $\beta<\nu$，且 $\nu<\mu$，，证明：关于 μ 几乎处处有 $\frac{\mathrm{d}\beta}{\mathrm{d}\mu}=\frac{\mathrm{d}\beta}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$．[提示：用习题 1.]
3. 当 $0\leqslant y\leqslant x\leqslant 1$ 时，设 $g(x,y)=2$；否则，$g=0$．设 μ 是 $\mathbb{R}^2$ 的测度，μ 关于勒贝格测度 λ^2 是绝对连续的且 $\mathrm{d}\mu/\mathrm{d}\lambda^2=g$．设从 $\mathbb{R}^2$ 映射到 $\mathbb{R}^1$ 的函数为 $T(x,y):=x$，且 τ 为像测度 $\mu\circ T^{-1}$．求 $\mathbb{R}$ 上的勒贝格测度 λ 关于 τ 的勒贝格分解及 Radon-Nikodym 导数 $\mathrm{d}\lambda_{ac}/\mathrm{d}\tau$．
4. 如果 μ，ν 是 σ-代数 $\mathcal{S}$ 上的有限测度，证明：ν 关于 μ 是绝对连续的当且仅当对每个 $\varepsilon>0$，存在 $\delta>0$，使得

对任意的 $A\in\mathcal{S}$，$\mu(A)<\delta$，蕴函着 $\nu(A)<\varepsilon$.

5. 勒贝格测度 λ 关于$[0,1]$上的计数测度 c 是绝对连续的，但不是 σ-有限的．证明：Radon-Nikodym 定理的结论在这种情况下不成立(这里没有具有 Radon-Nikodym 定理中 h 的性质的"导数"$d\lambda/dc$).
6. 对任意的测度空间$(X,\mathcal{S},\mu)$，证明：如果 $p<r<s$，$f\in\mathcal{L}^p(X,\mathcal{S},\mu)\cap\mathcal{L}^s(X,\mathcal{S},\mu)$，则$f\in\mathcal{L}^r(X,\mathcal{S},\mu)$. [提示：用 Rogers-Hölder 不等式．]
7. 回顾 ℓ^p 空间，ℓ^p 是空间 $\mathcal{L}^p(\mathbb{N},2^{\mathbb{N}},c)$，其中 c 是 $\mathbb{N}$ 上的计数测度．换句话说，$\ell^p:=\left\{\{x_n\}:\sum_{n\geq 0}|x_n|^p<\infty\right\}$，范数为$\left(\sum_{n\geq 0}|x_n|^p\right)^{1/p}$，$1\leqslant p<\infty$．证明：对 $0<p\leqslant r$，$\ell^p\subset\ell^r$(注意，在 L^p 有限测度的情况下，结论中相反的不等号成立．)

177

8. 考虑命题 4.2.1 的证明中定义的康托尔函数 g. 作为非递减函数，由定理 3.2.6 知，它定义了$[0,1]$上的一个测度 ν，满足对 $0\leqslant a<b\leqslant 1$，$\nu((a,b])=g(b)-g(a)$. 证明：$\nu\perp\lambda$，其中 λ 是勒贝格测度.
9. 在具有博雷尔 σ-代数$\mathcal{B}$的$[0,1]$中，令$\mu(A)=0$，对所有的 $A\in\mathcal{B}$，$\mu(A)=\infty$，对其他的所有 $A\in\mathcal{B}$. 证明：μ 是一个测度，如果 ν 是$\mathcal{B}$上关于μ 绝对连续的有限测度，则 $\nu\equiv 0$. [提示：参见 3.4 节习题 7～8. 因此，尽管 μ 不是 σ-有限的，Radon-Nikodym 定理的结论在 $h\equiv 0$ 时成立.]

5.6 符号测度

测度已经被定义为非负的，所以我们可以视它们为质量分布．"符号测度"将满足测度的定义，它除了有实值外，还可能有负值，如电荷的分布．

定义 给定一个可测空间$(X,\mathcal{S})$，符号测度(signed measure)是一个从$\mathcal{S}$映射到$[-\infty,\infty]$的可数可加函数μ，即$\mu(\varnothing)=0$，且对任意不相交的 $A_n\in\mathcal{S}$，$\mu\left(\bigcup_n A_n\right)=\sum_n\mu(A_n)$.

注： 定义中的和都假设为有定义的，以至于得不到$\mu(A)=+\infty$，$\mu(B)=-\infty$的集合 A，B. 如果 A，B 不相交，则结论显然成立．否则，$\mu(A\setminus B)+\mu(A\cap B)=+\infty$，$\mu(B\setminus A)+\mu(A\cap B)=-\infty$. 当这两项的和无有限时，至少同一个符号中有一个是无限的，另一个是无限的或是有限的．因此，$\mu(A\cap B)$是有限的，$\mu(A\setminus B)=+\infty$，$\mu(B\setminus A)=-\infty$. 但是$\mu(A\Delta B)=\mu(A\setminus B)+\mu(B\setminus A)$是无定义的，得到矛盾．因此，至多在$-\infty$和$+\infty$这两个值中有符号测度．实际上，符号测度空间通常由有限值的空间构成．

如果 ν，ρ 是两个测度，它们中至少一个是有限的，则 $\nu-\rho$ 是符号测度．下面的定理证明了所有的符号测度都是这种形式的，这里 ν，ρ 视为奇异的.

5.6.1 定理(哈恩-若尔当分解) 对可测空间$(X,\mathcal{S})$上的任意符号测度μ，有一个集合 $D\in\mathcal{S}$，使得对所有的 $E\in\mathcal{S}$，$\mu^+(E):=\mu(E\cap D)\geqslant 0$，$\mu^-(E):=-\mu(E\setminus D)\geqslant 0$. 那么，$\mu^+$，$\mu^-$都是测度，它们中至少一个是有限的，$\mu=\mu^+-\mu^-$，$\mu^+\perp\mu^-$($\mu^+$，$\mu^-$都是奇异的)．这些性质唯一地确定了$\mu^+$，$\mu^-$. 如果 F 是$\mathcal{S}$中另一个具有 D 的性质的任意集合，则$(\mu^++\mu^-)(F\Delta D)=0$. 对任意的 $E\in\mathcal{S}$，

178

$$\mu^+(E)=\sup\{\mu(G):G\subset E\},\quad \mu^-(E)=-\inf\{\mu(E):G\subset E\}.$$

证明 必要时用$-\mu$代替μ，对任意的 $E\in\mathcal{S}$，假设$\mu(E)>-\infty$. 令 $c:=\inf\{\mu(E):E\in\mathcal{S}\}$. 取 $c_n\downarrow c$，$E_n\in\mathcal{S}$且对所有的 n，满足 $\mu(E_n)<c_n$. 令 $A_1:=E_1$，递归地定义 $A_{n+1}:=A_n$ 当且仅当 $\mu(E_{n+1}\setminus A_n)\geqslant 0$. 否则，令 $A_{n+1}:=A_n\cup E_{n+1}$. 那么对所有的 n，$\mu(E_n\setminus A_n)\geqslant 0$，所以$\mu(E_n\cap A_n)\leqslant\mu(E_n)<c_n$. 现在，$A_1\subset A_2\subset\cdots$且$\mu(A_1)\geqslant\mu(A_2)\geqslant\cdots$. 令 $B_1:=\bigcup_n A_n$，则$\mu(B_1)<c_1$，$\inf\{\mu(E):E\subset B_1\}=c$. 那么通过另一种递归，存在一个递减的集合序列 B_n，且对所有的 n，$\mu(B_n)<c_n$. 令

$C:=\bigcap_n B_n$，则 $\mu(C)=c$. 因此，$c>-\infty$. 对任意的 $E\in\mathcal{S}$，$\mu(C\cap E)\leqslant 0$（否则，$\mu(C\setminus E)<c$）且 $\mu(E\setminus C)\geqslant 0$（否则，$\mu(C\cup E)<c$）. 令 $D:=X\setminus C$，则 D，μ^+，μ^- 都有所描述的性质.

对任意的 $E\in\mathcal{S}$，显然 $\mu^+(E)=\mu(E\cap D)\leqslant\sup\{\mu(G):G\subset E\}$. 对任意的 $G\subset E$，$\mu(G)=\mu^+(G)-\mu^-(G)\leqslant\mu^+(G)\leqslant\mu^+(E)$. 因此，$\mu^+(E)=\sup\{\mu(G):G\subset E\}$. 同样，$\mu^-(E)=-\inf\{\mu(G):G\subset E\}$. 为了证明 μ^+，μ^- 的唯一性，假设对非负奇异的测度 ρ 和 σ，有 $\mu=\rho-\sigma$. 那么对任意的可测集 $G\subset E$，$\rho(E)\geqslant\rho(G)\geqslant\mu(G)$. 因此 $\rho\geqslant\mu^+$. 否则，$\sigma\geqslant\mu^-$. 取 $H\in\mathcal{S}$，满足 $\rho(X\setminus H)=\sigma(H)=0$. 于是对任意的 $E\in\mathcal{S}$，$\rho(E)=\rho(E\cap H)=\mu(E\cap H)\leqslant\mu^+(E)$，所以 $\rho=\mu^+$，$\sigma=\mu^-$，这就证明了唯一性.

如果 F 是 $\mathcal{S}$ 中的另一个具有 D 的性质的集合，那么对任意的 $E\in\mathcal{S}$，

$$\mu^+(E)=\mu^+(E\cap D)=\mu(E\cap D)=\mu^+(E\cap F)=\mu(E\cap F).$$

所以 $\mu((D\Delta F)\cap E)=0$. 因此，$\mu^+(D\Delta F)=0=\mu(D\Delta F)=\mu^-(D\Delta F)$，证毕. □

X 分解为 D 和 C，这里 μ 分别是非负的和非正的，这种分解称为 X 关于 μ 的哈恩分解(Hahn decomposition). 等式 $\mu=\mu^+-\mu^-$ 称为 μ 的若尔当分解(Jordan decomposition). 测度 $|\mu|=\mu^++\mu^-$ 称为关于 μ 的全变差(total variation)测度. 两个符号测度 μ 和 ν 称为是奇异的当且仅当 $|\mu|$ 和 $|\nu|$ 是奇异的.

5.6.2 推论 *勒贝格分解和 Radon-Nikodym 定理(5.5.3 和 5.5.4)也成立，其中 ν 是一个有限的符号测度.*

179

证明 若尔当分解(5.6.1)给出 $\nu=\nu^+-\nu^-$，其中 ν^+，ν^- 是有限测度. 对 ν 应用定理 5.5.3，对 ν^+，ν^- 应用定理 5.5.4，然后相减便得到结果. □

例如，如果对 σ-代数中所有的 A，有 $\nu(A)=\int_A f\mathrm{d}\mu$，这里 μ 是 σ-有限测度，f 是可积函数，则 ν 是一个有限的符号测度，对于 μ 是绝对连续的. 对于 ν 的哈恩分解可以由集合 $D=\{x: f(x)\geqslant 0\}$ 来定义，且对任意的可测集 A，满足 $\nu^+(A)=\int_A f^+\mathrm{d}\mu$，$\nu^-(A)=\int_A f^-\mathrm{d}\mu$. 定义哈恩分解的另一个集合 D' 将满足 $\int_{D\Delta D'}|f|\,\mathrm{d}\mu=0$.

现在设 $\mathcal{A}$ 是集合 X 的子集的任意代数. 一个定义在 $\mathcal{A}$ 上的扩张的实值函数 μ 称为可数可加的，如果 $\mu(\varnothing)=0$ 且对 $\mathcal{A}$ 中的任意不相交的集合 A_1，A_2，…，使得只要 $A:=\bigcup_{n\geqslant 1}A_n$ 也在 $\mathcal{A}$ 中，就有 $\mu(A):=\sum_n\mu(A_n)$. 那么像对符号测度那样，μ 不能同时取 $-\infty$，$+\infty$ 这两个值. 我们称 μ 是有上界的，如果对某个 $M<+\infty$，使得对任意的 $A\in\mathcal{A}$，$\mu(A)\leqslant M$. 同样，μ 称为有下界的当且仅当 $-\mu$ 是有上界的. 回顾任意定义在一个代数上的可数可加的非负函数都有一个能扩张到由 $\mathcal{A}$ 生成的 σ-代数(定理 3.1.4.)上的可数可加函数，从而给出了一个测度. 为了使符号函测度也具有这个事实，则必须满足具有上界或下界.

5.6.3 定理 *如果 μ 是从代数 $\mathcal{A}$ 映射到 $[-\infty,\infty]$ 的可数可加测度，则 μ 可以扩张为由 $\mathcal{A}$ 生成的 σ-代数 $\mathcal{S}$ 上的一个符号测度，当且仅当 μ 是有上界的或下界的.*

证明 "必要性"：由哈恩-若尔当分解定理(5.6.1)知，对一个 σ-代数上的任意符号测度 μ，μ^+ 或 μ^- 是有限的. 因此，μ 有上界或下界. 反之，假设 μ 有上界或下界. 对任意的 $A\in\mathcal{A}$，设

$\mu^+(A):=\sup\{\mu(B): B\subset A\}$. 那么 $\mu^+\geqslant 0$(令 $B=\varnothing$)且 $\mu^+(\varnothing)=0$. 假设 $A_n\in\mathcal{A}$是不相交的, $A=\bigcup_n A_n\in\mathcal{A}$. 那么显然$\mu^+(A)\geqslant\sum_n\mu^+(A_n)$. 反之, 对任意的 $B\subset A$, $B\in\mathcal{A}$, $\mu(B)=\sum_{n\geqslant1}\mu(B\cap A_n)\leqslant\sum_{n\geqslant1}\mu^+(A_n)$. 所以$\mu^+(A)\leqslant\sum_{n\geqslant1}\mu^+(A_n)$. 因此, μ^+在$\mathcal{A}$上是可数可加的. 令$\mu^-:=(-\mu)^+$, 则可看出μ^-也是可数可加的且是非负的. 对每个 $A\in\mathcal{A}$, $\mu(A)=\mu^+(A)-\mu^-(A)$, 因此不可能假设$\mu^+(A)$和$\mu^-(A)$都是无限的. 那么由定理3.1.4, 每个μ^+和μ^-都可扩张为$\mathcal{S}$上的一个测度, 且至少有一个是有限的, 所以两个测度的差就是一个明确定义的扩张μ到$\mathcal{S}$的符号测度. □

180

下面的一个例子表明了为什么"上界或下界"可以从一个σ-代数上的可数可加性得到, 但不能从自身扩张到一个代数上.

*__5.6.4 命题__ *在一个代数$\mathcal{A}$上存在一个可数可加的实值函数, 它不能扩充为由$\mathcal{A}$生成的σ-代数$\mathcal{S}$上的一个符号测度.*

证明 设 X 是一个不可数集, 如$\mathbb{R}$. 记 $X=A\cup B$, 其中 A 和 B 是不相交的不可数集, 如 $A=(-\infty,0]$, $B=(0,\infty)$. 设$\mathcal{A}$是由所有的有限集及其补集组成的代数. 令 c 为计数测度. 对有限的 F, 令 $\mu(F):=c(A\cap F)-c(B\cap F)$. 对具有有限补 F 的 G, 令 $\mu(G):=-\mu(F)$.

如果 $C_n\in\mathcal{A}$, C_n 是不相交的, $C=\bigcup_n C_n$, 且 C 是有限的, 则所有的 C_n 是有限的, 且几乎有限多个都是空的, 所以显然可数可加性成立. 如果 C 有有限补, 那么 C_n 中恰有一个也有有限补, 记为 C_1. 于是对 $n>1$, C_n 是有限的, 且几乎有限多个都是空的. 我们有

$$X\setminus C_1=(X\setminus C)\cup\bigcup_{n\geqslant2}C_n$$

是一个不交并, 且 $X\setminus C_1$ 是有限的, 所以 $-\mu(C_1)=\mu(X\setminus C_1)=-\mu(C)+\sum_{n\geqslant2}\mu(C_n)$且 $\mu(C)=\sum_{n\geqslant1}\mu(C_n)$. 因此, μ 在代数$\mathcal{A}$上具有可数可加性. 因为 μ 没有上界或下界, 所以由定理 5.6.3 知, 它没有到$\mathcal{S}$上的可数可加性的扩张. □

对代数$\mathcal{A}$上的一个有限可加的实值函数 μ, 可数可加性和条件 $\mu(A_n)\to0$ 是等价的, 这里 $A_n\in\mathcal{A}$且 $A_n\downarrow\varnothing$, 正如在定理 3.1.1 中所证明的那样.

习题

1. 证明: 对命题 5.6.4 的证明中的代数$\mathcal{A}$, $\mathcal{A}$上的每个有限可加函数是可数可加的.
2. 对所有的实数 x, 设 $f(x):=x^2-6x+5$. 设 $\nu(A):=\int_A f(x)\,dx$, 对$\mathbb{R}$ 中的每个博雷尔集 A. 求$\mathbb{R}$ 关于 ν 的哈恩分解和关于 ν 的若尔当分解.

181

3. 对什么样的多项式 P, 习题 2 中 f 可由 P 来代替并且给出一个在$\mathbb{R}$ 的博雷尔集上有明确定义的符号测度?
4. 拓扑空间$\mathcal{S}$上的拉东测度 μ 定义为在每个具有紧闭包的博雷尔集上的一个实值函数, 且使得对每个固定的紧集 K, μ 到 K 的博雷尔子集的限制是可数可加的. 证明: 对$\mathbb{R}$ 上的每个连续函数 f, $\mu(A):=\int_A f(x)\,d\lambda(x)$ 定义了一个若尔当测度. 注意: 拉东测度不是一个测度, 除非$\mathcal{S}$是紧的. 一些作者也可能在定义拉东测度时加强条件, 如正则性, 见 7.1 节.
5. 证明: $\mathbb{R}$ 上的任意拉东测度 μ 有一个若尔当分解 $\mu=\mu^+-\mu^-$, 其中 μ^+ 和 μ^- 是$\mathbb{R}$ 上的非负、σ-有限的拉东测度, 且可以唯一地扩张为所有博雷尔集的 σ-代数上的通常的非负测度.
6. 对$\mathbb{R}$ 上的拉东测度 $\mu(A)=\int_A x^3-x\,dx$ 进行习题 5 中的分解.

7. 设 $\mu(A):=\int_A f(x)\mathrm{d}\lambda(x)$．证明：$\mu$ 是$\mathbb{R}$上的一个拉东测度，但 μ 不能扩张为$\mathbb{R}$上的一个符号测度.
8. 设 μ 和 ν 是有限测度，使得 ν 关于 μ 是绝对连续的，且 $f=\mathrm{d}\nu/\mathrm{d}\mu$. 证明：对每个 $r>0$，$\nu-r\mu$ 的哈恩分解给出集合 A 和 A^c，使得在 A 上，$f\geqslant r$ a. e.，在 A^c 上，$f\leqslant r$ a. e..
9. 对有限测度，由哈恩分解定理证明：Radon-Nikodym 定理．[提示：参考习题 8 来定义 f.]
10. 拓扑空间上的一个实值函数 f 的支集(support)是闭集 $\{x: f(x)\neq 0\}$ 的闭包. 设$\mathcal{L}$是$\mathbb{R}$上具有紧支集的连续实值函数的集合.

(a)证明：对任意的 $f\in\mathcal{L}$和若尔当测度 μ，$\int f\mathrm{d}\mu$ 是有 定义的.

(b)证明：$\mathcal{L}$是一个在 4.5 节中所定义的斯通向量格点.

(c)证明：如果 I 是从$\mathcal{L}$映射到$\mathbb{R}$ 的线性函数且 $I(f)\geqslant 0$，其中 $f\geqslant 0$，那么 L 是一个准整数.

注释

5.1 节　Cauchy(1821，p. 373)对有限和证明了不等式 5.1.4，换句话说，μ 是一个有限集上的计数测度．Bunyakovsky(1859)对黎曼积分证明了不等式 5.1.4. H. A. Schwarz(1885，p. 351)较晚才证明此不等式．此不等式普遍被称为“施瓦茨不等式”，近来，被称为“柯西–施瓦茨不等式”．不等式 5.1.2 一般被称为“赫尔德不等式”，但 Hölder(1889)声称他是受到 L. J. Rogers(1888)的论文的启发．习题 9 将在他们论文发现的结果 I 和不等式 5.1.2 联系起来．Rogers(1888，p. 149，§3，(1)和(4))证明了习题 9 中的“Rogers 不等式”，对有限和及积分 $\int_a^b \mathrm{d}x$. Hölder(1889，p. 44)证明了习题 9 中的对有限和的“过去的赫尔德不等式”. F. Riesz(1910，p. 456)证明了积分形式的不等式．对积分的流形式，Rogers 做了一些其他研究，参见 Stanley(1971)．赫尔德也由于对一个函数 f 上的“赫尔德条件”(即 $|f(x)-f(y)|\leqslant K|x-y|^{\alpha}$，$0<\alpha\leqslant 1$)及其他，包括有限群论中的一些著作而闻名，参见 van der Waerden(1939)．不等式 5.1.5 一般被称为“Minkowski 不等式”. H. Minkowski(1907，p. 95)证明了有限和的情况．关于 Minkowski 的生平，有一部由他的长女所著的论文集：Rüdeberg(1973)．F. Riesz(1910，p. 456)将不等式扩张到积分的形式．以上的部分注释是以 Hardy、Littlewod 和 Polya(1952)为基础的，Hardy、Littlewod 和 Polya 很谦虚地写道“他们从来没承担过系统的著书目录的研究”. Hardy 和 Littlewod 是他们那个时期英国数学界的领军人物，成果最丰富的数学合作者之一，一起写过 100 多篇论文．[182]

5.2 节　F. Riesz(1909，1910)定义了$\mathcal{L}^p$空间(对 $1<p<2$ 和 $2<p<\infty$)，并证明了它们的完备性．Riesz(1906)定义了 L^2 距离．E. Fischer(1907)证明了 L^2 的完备性．关于“Riesz-Fischer”定理的相关内容，参见 5.4 节及其注释．Banach(1982)定义了线性赋范空间和巴拿赫空间(他称之为“espaces de type B”)．Wiener(1922)也单独定义了线性赋范空间．Banach(1932)完善了这类空间的理论，使得这类空间用他的名字来命名．Steinhaus(1961)描述巴拿赫“他一出生，就被送给一个叫 Banachowa 的清洁工来抚养，她住在雅典……在巴拿赫 15 岁时，他不得不通过代课来谋生．”

5.3 节　希尔伯特在 1904 ~ 1910 年出版的论文中，发展了“希尔伯特空间”方法，用于解决积分方程．线性(完备)内积空间，或希尔伯特空间的定义含蓄地表达在希尔伯特的著作中，由 E. Schmidt(1907，1908)给出了明确的表达和扩展，如同“欧几里得空间”，施密特强调希尔伯特空间的几何观点．正交分解定理(5.3.8)的证明由 F. Riesz(1934)给出，虽然定理本身很久远了．Riesz 关于这种思想对 Levi(1906，§7)表达感激.

为了证明$\mathbb{R}^2$中的这个距离，例如，这样定义距离

$$d((x,y),(u,v)) := ((x-u)^2+(y-v)^2)^{1/2},$$

它在坐标旋转时是不变的，可用三角恒等式$\cos^2\theta+\sin^2\theta=1$，此公式依赖于平面几何中经典的毕达哥拉斯定理(也称为勾股定理). 在这种意义下，它是圆形的，由此断言定理 5.3.6 事实上证明了毕达哥拉斯定理. 通过《欧几里得几何原本》，经典的毕达哥拉斯定理早已为人所知：欧几里得在大约公元前 300 年很出名，早期的希腊几何书仅有几本还保留着. 毕达哥拉斯(大约 560B. C. -480B. C.)创立了一个流派或秘密社会学，在其他的事情中，对数学很感兴趣. "mathematics"这几个字母就来源于希腊语 $\mu\alpha\theta\eta\alpha\tau\iota\kappa o\iota$，在毕达哥拉斯学派中那些人的地位得到改善. O. Neugebauer(1935, I, p. 180; 1957, p. 36)翻译了巴比伦的楔形文章，证明了"毕达哥拉斯定理"，或至少有它的许多例子，这些例子从汉摩拉比时期(即公元前 1600 以前)就已经为众人所知，参见 Buck(1980). 但不清楚的是谁首先证明了这个定理.

5.4 节 存在一个没有规范正交基(不完备、不可数维数)的内积空间(Dixmier, 1953). 研究的第一个无限规范正交集就是出现在傅里叶级数中的三角函数. 对通常长度的测度(勒贝格测度)，函

183

数列 $\cos(kx)$ $(k=0, 1, \cdots)$ 在$\mathcal{L}^2([0, \pi])$中是正交的. Parseval(1799, 1801)在特殊的情况下也有相同的结论(定理 5.4.4)，在讨论这些三角函数的正交时，对 $k=0$ 与 $k=1, 2\cdots$ 要求是不同的数值因子. 帕塞瓦尔引用了 Euler(1755)的一些思想. F. W. Bessel(1784—1846)在与"贝塞尔函数"相关的数学领域中是众所周知的，其包含了极坐标中半径为 r 的函数 $J_n(r)$(拉普拉斯算子的特征函数)，

$$((\partial^2/\partial x^2)+(\partial^2/\partial y^2))J_n(r) = c_nJ_n(r), \quad c_n\in\mathbb{R}$$

和相关函数. 关于贝塞尔函数，除了很多书表以外，至少有 7 本专著，例如，Watson(1944). 贝塞尔主要从事天文学方面的工作(他的著作收集在 Bessel, 1875). 在其他的成就中，除了 Fricke (1970, p. 100)计算太阳的距离外，他是第一个计算恒星的距离的人. Riesz(1907a)称贝塞尔已经陈述了他的有关连续函数的不等式.

"格拉姆–施密特"规范正交化(定理 5.4.6)首先由丹麦统计学家 J. P. Gram(1879; 1883, p. 48)发现，在 Schweder(1980, p. 118)中给出了著作的注解. 规范正交化经过 E. Schmidt(1907, p. 442)的扩张更为人们所熟知. 规范正交基理论在 1900 年后得到快速发展，在 Hurwitz(1903)和 Fatou (1906)的论文和其他著作中，有三角级数的完善论述. 希尔伯特和施密特(见前部分的注)都用到了规范正交基. Riesz-Fischer 定理(5.4.5)是以 Fisher(1907)的论文(关于 L^2 的完备性)，特别是 Riesz(1907a)为基础来命名的. 有关其历史参见 Siegmund-Schultze(1982, Kap. 5). 正交分解定理(5.3.8)也可用如下的方法来证明：给定 $x\in H$ 和子空间 F 的规范正交基$\{e_\alpha\}$，令 $y:=\sum_\alpha (x,e_\alpha)e_\alpha$. 定理首先证明当 F 是可分的情况，其中格拉姆–施密特过程(5.4.6)给出了 F 的一组规范正交基.

5.5 节 Riesz-Fréchet 定理(5.5.1)一般称为"Riesz 表示定理"，由分别注释在同一期的《Comptes Rendus》中的 Riesz(1907b)和 Fréchet(1907)对此进行了论述. Radon-Nikodym 定理应归功于 Radon(1913, p. 1342—1351). Daniell(1920)和 Nikodym(1930). 勒贝格分解和 Radon-Nikodym 定理的联合证明是由 von Neumann(1940)给出的.

当 μ 不是 σ-有限时，习题 5 和 9 表明了 Radon-Nikodym 定理不成立，但不总是不成立的，参见 Bell 和 Hagood(1981).

5.6 节 对定义在区间$[a, b]$上的一个实值函数 f，f 的全变差定义为所有和

$$\sum_{1\leq j\leq n} | f(x_j) - f(x_j - 1) |$$

的上确界，其中 $a \leq x_0 \leq x_1 \leq \cdots \leq x_n \leq b$. 称一个函数在$[a, b]$上是有界变差的，当且仅当它是有限的全变差 . Jordan(1881)定义了有界变差函数且证明了每个这样的函数是两个非递减函数的差，记为 $f = g - h$. 在更一般的情形下，"若尔当分解"是为纪念此结果而命名的，并且可以考虑它的一个扩张 . 如果 μ 是$[a, b]$上的一个有限符号测度，且 $f(x) = \mu([a, x])$，则对 $a \leq x \leq b$，我们可取 $g(x) = \mu^+([a, x])$ 和 $h(x) = \mu^-([a, x])$. [184]

若尔当对数学的其他许多分支起着重要的辅助作用，包括拓扑和有限群论 .（见他的 Oeuvres，下面有参考文献 .）勒贝格(1910, p. 381—382)定义了符号测度并研究了形式为

$$\mu(A) = \int_A f(x)\,\mathrm{d}m(x)$$

的测度，其中对一个测度 m，$f \in \mathcal{L}^1(m)$. 那么 μ^+ 和 μ^- 由 f^+ 和 f^- 这样的"不定积分"给出 . 当前的"哈恩分解"和若尔当分解是由 Hahn(1921, p. 393—406)证明的 . 尽管哈恩称他 1921 的书是第一卷，但第二卷在 1934 年他去世前还没出版 . 最终，他的大部分著作以 Hahn 和 Rosenthal(1948)的形式出版了 . 关于 Arthur Rosenthal(1887—1959)，参见 Haupt(1960). [185]

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到 .

Banach, Stefan (1922). Sur les opérations dans les ensembles abstraits et leurs applications aux équations intégrales. *Fundamenta Math. 3*: 133–181.

——— (1932). *Théorie des operations linéaires*. Monografje Matematyczne, Warsaw. 2d ed. (posth.), Chelsea, New York, 1963.

Bell, W. C., and Hagood, J. W. (1981). The necessity of sigma-finiteness in the Radon-Nikodym theorem. *Mathematika 28*: 99–101.

*Bessel, Friedrich Wilhelm (1875). *Abhandlungen*. Ed. Rudolf Engelmann. 3 vols. Leipzig.

Buck, R. Creighton (1980). Sherlock Holmes in Babylon. *Amer. Math. Monthly 87*: 335–345.

*Bunyakovsky, Viktor Yakovlevich (1859). Sur quelques inégalités concernant les intégrales ordinaires et les intégrales aux différences finies. *Mémoires de l'Acad. de St.-Petersbourg* (Ser. 7) *1*, no. 9.

Cauchy, Augustin Louis (1821). *Cours d'analyse de l'École Royale Polytechnique* (Paris). Also in *Oeuvres complètes d'Augustin Cauchy* (Ser. 2) *3*. Gauthier-Villars, Paris (1897).

Daniell, Percy J. (1920). Stieltjes derivatives. *Bull. Amer. Math. Soc. 26*: 444–448.

Dixmier, Jacques (1953). Sur les bases orthonormales dans les espaces préhilbertiens. *Acta Sci. Math. Szeged 15*: 29–30.

Euler, Leonhard (1755). *Institutiones calculi differentialis*. Acad. Imp. Sci. Petropolitanae, St. Petersburg; also in *Opera Omnia* (Ser. 1) *10*.

Fatou, Pierre (1906). Séries trigonométriques et séries de Taylor. *Acta Math. 30*: 335–400.

Fischer, Ernst (1907). Sur la convergence en moyenne. *Comptes Rendus Acad. Sci. Paris 144*: 1022–1024.

Fréchet, Maurice (1907). Sur les ensembles de fonctions et les opérations linéaires. *Comptes Rendus Acad. Sci. Paris 144*: 1414–1416.

Fricke, Walter (1970). Bessel, Friedrich Wilhelm. *Dictionary of Scientific Biography, 2*, pp. 97–102.

*Gram, Jørgen Pedersen (1879). *Om Räkkeudviklinger, bestemte ved Hjälp of de mindste Kvadraters Methode* (Doctordissertation). Höst, Copenhagen.

——— (1883). Uber die Entwickelung reeller Functionen in Reihen mittelst der Methode der Kleinsten Quadrate. *Journal f. reine u. angew. Math. 94*: 41–73.

Hahn, Hans (1921). *Theorie der reellen Funktionen*, "*I. Band*." Julius Springer, Berlin.

——— (posth.) and Arthur Rosenthal (1948). *Set Functions*. Univ. New Mexico Press.

Hardy, Godfrey Harold, John Edensor Littlewood, and George Polya (1952). *Inequalities*. 2d ed. Cambridge Univ. Press. Repr. 1967.

Haupt, O. (1960). Arthur Rosenthal (in German). *Jahresb. deutsche Math.-Vereinig. 63*: 89–96.

Hilbert, David (1904–1910). Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. *Nachr. Ges. Wiss. Göttingen Math.-Phys. Kl. 1904*: 49–91, 213–259; *1905*: 307–338; *1906*: 157–227, 439–480; *1910*: 355–417. Also published as a book by Teubner, Leipzig, 1912, repr. Chelsea, New York. 1952.

Hölder, Otto (1889). Über einen Mittelwerthssatz. *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. 1889*: 38–47.

Hurwitz, A. (1903). Über die Fourierschen Konstanten integrierbarer Funktionen. *Math. Annalen 57*: 425–446.

Jordan, Camille (1881). Sur la série de Fourier. *Comptes Rendus Acad. Sci. Paris 92*: 228–230. Also in Jordan (1961–1964), *4*, pp. 393–395.

——— (1961–1964). *Oeuvres de Camille Jordan*. J. Dieudonné and R. Garnier, eds. 4 vols. Gauthier-Villars, Paris.

Lebesgue, Henri (1910). Sur l'intégration des fonctions discontinues. *Ann. Scient. Ecole Normale Sup.* (Ser. 3) *27*: 361–450. Also in Lebesgue (1972–1973) *2*, pp. 185–274.

——— (1972–1973). *Oeuvres scientifiques*. 5 vols. L'Enseignement Mathématique. Institut de Mathématique, Univ. Genève.

Levi, Beppo (1906). Sul principio de Dirichlet. *Rendiconti Circ. Mat. Palermo 22*: 293–360.

Minkowski, Hermann (1907). *Diophantische Approximationen*. Teubner, Leipzig.

——— (1973, posth.). *Briefe an David Hilbert*. Ed. Lily Rüdenberg and Hans Zassenhaus. Springer, Berlin.

Neugebauer, Otto (1935). *Mathematische Keilschrift-Texte*. 2 vols. Springer, Berlin.

——— (1957). *The Exact Sciences in Antiquity*. 2d ed. Brown Univ. Press. Repr. Dover, New York (1969).

von Neumann, John [Johann] (1940). On rings of operators, III. *Ann. Math. 41*: 94–161.

Nikodym, Otton Martin (1930). Sur une généralisation des mesures de M. J. Radon. *Fundamenta Math. 15*: 131–179.

*Parseval des Chênes, Marc-Antoine (1799). Mémoire sur les séries et sur l'intégration complète d'une équation aux différences partielles linéaires du second ordre, à coefficiens constans. *Mémoires présentés à l'Institut des Sciences, Lettres et Arts, par divers savans, et lus dans ses assemblées. Sciences math. et phys. (savans étrangers) 1* (1806): 638–648.

*——— (1801). Intégration générale et complète des équations de la propagation du son, l'air étant considéré avec ses trois dimensions. Ibid., pp. 379–398.

*Radon, Johann (1913). Theorie und Anwendungen der absolut additiven Mengenfunktionen. *Sitzungsber. Akad. Wiss. Wien Abt. IIa*: 1295–1438.

Riesz, Frédéric [Frigyes] (1906). Sur les ensembles de fonctions. *Comptes Rendus Acad. Sci. Paris 143*: 738–741.

——— (1907a). Sur les systèmes orthogonaux de fonctions. *Comptes Rendus Acad. Sci. Paris 144*: 615–619.

——— (1907b). Sur une espèce de géométrie analytique des fonctions sommables. *Comptes Rendus Acad. Sci. Paris 144*: 1409–1411.

——— (1909). Sur les suites de fonctions mesurables. *Comptes Rendus Acad. Sci. Paris 148*: 1303–1305.

——— (1910). Untersuchungen über Systeme integrierbarer Funktionen. *Math. Annalen 69*: 449–497.

——— (1934). Zur Theorie des Hilbertschen Raumes. *Acta Sci. Math. Szeged 7*: 34–38.

Rogers, Leonard James (1888). An extension of a certain theorem in inequalities. *Messenger of Math. 17*: 145–150.

Rüdenberg, Lily (1973). Erinnerungen an H. Minkowski. In Minkowski (1973), pp. 9–16.

Schmidt, Erhard (1907). Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. *Math. Annalen 63*: 433–476.

——— (1907). Auflösung der allgemeinen linearen Integralgleichung. *Math. Annalen 64*: 161–174.

——— (1908). Über die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten. *Rend. Circ. Mat. Palermo 25*: 53–77.

*Schwarz, Hermann Amandus (1885). Über ein die Flächen kleinsten Flächeninhalts betreffendes Problem der Variationsrechnung. *Acta Soc. Scient. Fenn. 15*: 315–362.

Schweder, Tore (1980). Scandinavian statistics, some early lines of development. *Scand. J. Statist 7*: 113–129.

Siegmund-Schultze, Reinhard (1982). Die Anfänge der Funktionalanalysis und ihr Platz im Umwälzungsprozess der Mathematik um 1900. *Arch. Hist. Exact Sci. 26*: 13–71.

Stanley, Richard P. (1971). Theory and application of plane partitions. *Studies in Applied Math. 50*: 167–188.

Steinhaus, Hugo (1961). Stefan Banach, 1892–1945. *Scripta Math. 26*: 93–100.

van der Waerden, Bartel Leendert (1939). Nachruf auf Otto Hölder. *Math. Ann. 116*: 157–165.

Watson, George Neville (1944). *A Treatise on the Theory of Bessel Functions*, 2d ed. Cambridge Univ. Press, repr. 1966; 1st ed., 1922.

Wiener, Norbert (1922). Limit in terms of continuous transformation. *Bull. Soc. Math. France 50*: 119–134.

第6章　范数空间的凸集和对偶性

泛函分析主要研究无穷维线性空间，例如，巴拿赫空间和希尔伯特空间，它们大多由函数或函数等价类组成. 每个巴拿赫空间 X 有一个对偶空间 X'，X'定义为由 X 映射到域$\mathbb{R}$ 或$\mathbb{C}$ 的所有连续线性函数的集合.

对偶的一个主要例子是 L^p 空间. 设$(X, \mathcal{S}, \mu)$是一个测度空间. 令 $1<p<\infty$，$1/p+1/q=1$. 那么对 L^p 中的 f 及 L^q 中的 g，通过线性函数 $f\mapsto \int fg\mathrm{d}\mu$ 可以证明 L^p 和 L^q 是互为对偶的. 对 $p=q=2$，L^2 是希尔伯特空间，前面已经证明了：希尔伯特空间上的任意连续线性函数是由与 H 中的一个固定元素的内积给出的(定理5.5.1).

除了线性子空间，向量空间 S 中最自然的和应用最频繁的子空间就是凸子集 C，使得对 C 中的任意 x，y 及 $0<t<1$，有 $tx+(1-t)y\in C$. 这些集合在6.2节和6.6节中进行讨论. 一个函数图形上方的区域是凸的就称这个函数为凸函数. 6.3节讨论了凸函数. 凸集和凸函数是现代实函分析中主要的研究课题.

6.1　利普希茨函数、连续函数及有界函数

设(S, d)及(T, e)是度量空间. 从 S 映射到 T 的函数f称为利普希茨的或利普希茨，当且仅当对某个 $K<\infty$，

$$e(f(x),\ f(y))\leqslant Kd(x,y),\quad \forall x,y\in S.$$

那么记 $\|f\|_L$ 为最小的 K(K 是存在的). 通常 $T=\mathbb{R}$，且有通常的度量.

回顾一个度量(或赋范)空间的闭子集上的连续实值函数通过扩张定理2.6.4可以扩张到整个空间. 一个非闭子集上的连续函数不一定能扩张，例如，开区间$(0, 1)$上的函数 $x\mapsto\sin(1/x)$，不能扩张为在0点连续. 对一个利普希茨函数，从任意子集上扩张都是可能的.

188

6.1.1 定理　*设(S, d)是任意度量空间，E 是 S 的任意子集，f 是 E 上的任意实值利普希茨函数. 那么 f 可以扩张到 S 且没有递增的 $\|f\|_L$.*

证明　令 $M:=\|f\|_L$. 假设我们有一个函数f_α的包含链，f_α是从 S 的子集 E_α 映射到$\mathbb{R}$ 的函数，且对所有的 α，$\|f_\alpha\|_L\leqslant M$. 设 g 是f_α的并，则 g 是一个满足 $\|g\|_L\leqslant M$ 的函数. 因此，由佐恩引理，它能够将f扩张到一个额外的点 $x\in S\setminus E$ 上.

实数 y 是$f(x)$的一个可能值，当且仅当对所有的 $u\in E$，$|y-f(u)|\leqslant Md(u, x)$，或等价于下面的两个条件成立：

(i) $-Md(u, x)\leqslant y-f(u)$，$\forall u\in E$，

(ii) $y-f(v)\leqslant Md(x, v)$，$\forall v\in E$.

这样的 y 存在，当且仅当

$$\sup_{u\in E}(f(u)-Md(u,x))\leqslant\inf_{v\in E}(f(v)+Md(v,x)).\qquad (*)$$

现在由假设，对所有的 u，$v\in E$，

$$f(u)-f(v)\leqslant Md(u,v)\leqslant Md(u,x)+Md(x,v),$$

$$f(u) - Md(u,x) \leqslant f(v) + Md(x,v).$$

这意味着(*)式成立，所以扩张到 x 是可能的. □

例如，如果 E 不是 S 的一个闭子集，令 $\{x_n\}$ 是 E 中收敛到 $S \setminus E$ 中一点 x 的任意序列. 那么 $\{f(x_n)\}$ 是一个收敛到某个实数的柯西序列，这个实数可以定义为 $f(x)$，因为它不依赖于特殊序列 $\{x_n\}$ 且 $f(x)$ 的选择是唯一的. 如果 x 不在 E 的闭包中，则 $f(x)$ 的选择可能不是唯一的(见习题 2).

设$(X, |\cdot|)$和$(Y, \|\cdot\|)$是两个线性赋范空间(如 5.2 节中所定义的). 一个从 X 映射到 Y 的线性函数 T，通常称为是一个算子(operator)，这个算子是有界的当且仅当它是利普希茨的.(注：通常一个非线性函数称为有界的，如果它的值域是有界的，但是一个线性函数 T 的值域是一个有界集，当且仅当 $T\equiv 0$.)

189

6.1.2 定理 给定一个从 X 映射到 Y 的线性算子 T. 其中$(X, |\cdot|)$和$(Y, \|\cdot\|)$是线性赋范空间，下面的结论是等价的：

(I) T 是有界的.

(II) T 是连续的.

(III) T 在 0 点是连续的.

(IV) T 在 $\{x\in X: |x| \leqslant 1\}$ 上的值域是有界的.

证明 每个利普希茨函数(线性或非线性的)是连续的，故(I)推出了(II)，进而就推出了(III). 如果(III)成立，取 $\delta>0$，使得 $|x| \leqslant \delta$，就推出了 $\|Tx\| \leqslant 1$. 那么对 X 中任意的 x 且 $0<|x| \leqslant 1$, $\|Tx\| = \|(|x|/\delta)T(\delta x/|x|)\| \leqslant |x|/\delta \leqslant 1/\delta$ 且 $\|T0\| =0$，故(IV)成立.

现在，如果(IV)成立，即当 $|x| \leqslant 1$ 时，$\|Tx\| \leqslant M$，则对所有的 $x\neq u\in X$，$\|Tx-Tu\| = \| |x-u| T((x-u)/|x-u|)\| \leqslant M|x-u|$，故(I)成立，这就完成了证明. □

例如，令 H 是希尔伯特空间 $L^2(X, \mathcal{S}, \mu)$. 设 M 是 $\mathcal{L}^2(X\times X, \mu\times\mu)$ 中的一个函数. 对每个 $f\in H$，令 $T(f)(x) := \int M(x,y)f(y)\,d\mu(y)$. 由 Tonelli-Fubini 定理和 Cauchy-Bunyakovsky-Schwarz 不等式知，对几乎所有的 x，$T(f)(x)$是明确定义的且

$$\int T(f)(x)^2 d\mu(x) \leqslant \int\left(\int M(x,y)^2 d\mu(y)\right)\int f(t)^2 d\mu(t)\, d\mu(x)$$

$$\leqslant \int f^2 d\mu \int M^2 d(\mu\times\mu) < \infty,$$

故 T 是有界线性算子.

设 K 为$\mathbb{R}$或$\mathbb{C}$. 定义在线性空间 X 上的一个函数 f 在 K 上称为是线性的，如果对 K 中所有的 c 和 X 中所有的 x, y，有 $f(cx+y)=cf(x)+f(y)$. $\mathbb{R}$ 上的线性函数称为是实线性的，$\mathbb{C}$ 上的函数称为是复线性的. 对 K 上的任意线性赋范空间$(X, \|\cdot\|)$，记 X' 为从 X 映射到 K 的所有线性连续函数(通常称为泛函或型)的集合. 对每个 $f\in X'$，令

$$\|f\|' := \sup\{|f(x)|/\|x\| : 0\neq x\in X\},$$

则由定理 6.1.2 知，$\|f\|'<\infty$ 且$(X', \|\cdot\|')$称为线性赋范$(X, \|\cdot\|)$的对偶(dual)或对偶空间(dual space).

190

现在，如果$(X, \mathcal{S}, \mu)$是一个测度空间，$1<p<\infty$ 且 $q=p/(p-1)$，则可测函数 f 的等价类的空间 $L^p(X, \mathcal{S}, \mu)$(由定理 5.1.5 和定理 5.1.2)是一个巴拿赫空间，且 f 满足 $\|f\|_p^p := \int |f|^p d\mu < \infty$，且范

数为 $\|\cdot\|_p$，对每个 $g\in\mathcal{L}^{p'}$，由 Rogers-Hölder 不等式 $f\mapsto\int fg\mathrm{d}\mu$ 是 L^p 上的一个有界线性型. 换句话说，这个线性型属于 $(L^p)'$（在 6.4 节中证明了 $(L^p)'$ 的所有元素都是由这种方式产生的.）

6.1.3 定理 对 $K=\mathbb{R}$ 或 $\mathbb{C}$ 上的任意线性赋范空间 $(X,\|\cdot\|)$，$(X',\|\cdot\|')$ 是一个巴拿赫空间.

证明 显然，X' 是 K 上的一个线性空间. 对任意的 $f\in X'$ 及 $c\in K$，$\|cf\|'=|c|\|f\|'$. 如果 $g\in X'$，则对所有的 $x\in X$，$|(f+g)(x)|\le|f(x)|+|g(x)|$，故 $\|f+g\|'\le\|f\|'+\|g\|'$. 因此 $\|\cdot\|'$ 是一个半范数. 如果在 X' 中 $f\neq0$，则对 X 中的某个 x，有 $f(x)\neq0$，故 $\|f\|'>0$ 且 $\|\cdot\|'$ 是一个范数. 设 $\{f_n\}$ 是一个柯西序列. 那么对每个 $x\in X$，

$$|f_n(x)-f_m(x)|\le\|f_n-f_m\|'\|x\|,$$

则 $\{f_n(x)\}$ 是 K 中的柯西序列且收敛到某个 $f(x)$. 作为一个线性函数的逐点极限，f 是线性的. 现在 $\{f_n\}$ 作为一个柯西序列是有界的，所以对于某个 $M<\infty$，对所有的 n，有 $\|f_n\|'\le M$. 于是对于所有的 $x\in X$ 且 $x\neq0$，有

$$|f(x)|/\|x\|=\lim_{n\to\infty}|f_n(x)|/\|x\|\le M,$$

所以 $f\in X'$ 且 $\|f\|'\le M$. 同样，对于任意的 m，

$$\|f_m-f\|'\le\limsup_{n\to\infty}\|f_m-f_n\|'\to0,\quad m\to\infty,$$

所以柯西序列收敛到 f 且 X' 是完备的. □

对一个赋范空间上的线性有界形式，下面的扩张定理对对偶定理至关重要且是泛函分析中三或四个著名定理之一.

6.1.4 定理（哈恩-巴拿赫定理） 设 $(X,\|\cdot\|)$ 是 $K=\mathbb{R}$ 或 $\mathbb{C}$ 上的任意线性赋范空间，E 是任意的线性子空间（与 X 上的范数一样），且 $f\in E'$. 那么 f 可以扩张为 X' 中的一个元素，且有同样的范数 $\|f\|'$.

证明 设在 E 上 $M:=\|f\|'$. 则由定理 6.1.2 的证明知，f 在 E 上是利普希茨的且 $\|f\|_L\le M$. 首先假设 $K=\mathbb{R}$，则由定理 6.1.1 知，对任意的 $x\in X\setminus E$，f 可以扩张为 $E\cup\{x\}$ 上的一个利普希茨函数且没有递增的 $\|f\|_L$. 包含 $E\cup\{x\}$ 的最小线性子空间 F 中的每个元素可以写为 $u+cx$，对某个唯一的 $u\in E$ 及 $c\in\mathbb{R}$. 设 $f(u+cx):=f(u)+cf(x)$，则 f 是线性的，且如果 $c=0$，$|f(u+cx)|\le M\|u+cx\|$ 成立，故假设 $c\neq0$. 于是

[191]

$$|f(u+cx)|=\left|-cf\left(-\frac{u}{c}-x\right)\right|\le|c|M\left\|-\frac{u}{c}-x\right\|=M\|u+cx\|.$$

故在 F 上 $\|f\|'\le M$. G 上的一系列 f 扩张为线性子空间上的线性函数且 $\|G\|'\le M$，其并也是具有同样性质的一个扩张. 因此，由佐恩引理，在定理 6.1.1 的证明中，对 $K=\mathbb{R}$ 定理已经证明了.

如果 $K=\mathbb{C}$，设 g 是 f 的实部. 那么 g 是一个实线性型，在 E 上 $\|g\|'\le\|f\|'$. 设 h 是 f 的虚部，所以 h 是实值的，是 E 上的实线性型且对所有的 $u\in E$，$f(u)=g(u)+\mathrm{i}h(u)$. 于是 $f(\mathrm{i}u)=g(\mathrm{i}u)+\mathrm{i}h(\mathrm{i}u)=\mathrm{i}f(u)=\mathrm{i}g(u)-h(u)$，故对所有的 $u\in E$，$h(u)=-g(\mathrm{i}u)$，$f(u)=g(u)-\mathrm{i}g(\mathrm{i}u)$. 通过实数情况，将 g 扩张为所有 X 上的一个实线性型且 $\|g\|'\le M$，对所有的 $x\in X$，定义 $f(x):=g(x)-\mathrm{i}g(\mathrm{i}x)$，则由于 g 是线性的，故 f 是 $\mathbb{R}$ 上的线性函数，且对任意的 x，

$$f(\mathrm{i}x)=g(\mathrm{i}x)-\mathrm{i}g(-x)=\mathrm{i}g(x)+g(\mathrm{i}x)=\mathrm{i}f(x),$$

因此 f 是复线性的. 对任意的 $x\in E$ 且 $f(x)\neq0$，令 $\gamma:=(g(x)+\mathrm{i}g(\mathrm{i}x))/|f(x)|$，则 $|\gamma|=1$，$f(\gamma x)\ge0$，于是 $|f(x)|=|f(\gamma x)/\gamma|=|f(\gamma x)|=f(\gamma x)=g(\gamma x)\le M\|\gamma x\|=M\|x\|$，所以

$\|f\|'\leq M$，完成了证明. □

对任意的线性赋范空间$(X,\ \|\cdot\|)$，有从 X 到 $X'':=(X')'$的一个自然映射 I''，对每个 $x\in X$，$f\in X'$，定义 $I''(x)(f):=f(x)$. 显然，I''是线性的. 令 $\|\cdot\|''$是 X''上的范数，其中 X''是 X'的对偶. 由$\|\cdot\|'$的定义得，对所有的 $x\in X$，$\|I''x\|''\leqslant x$.

6.1.5 推论 对任意的赋范空间$(X,\ \|\cdot\|)$和任意的 $x\in X$ 且 $x\neq 0$，有一个 $f\in X'$且 $\|f\|'=1$，$f(x)=\|x\|$. 因此对所有的 $x\in X$，$\|I''x\|''=x$.

证明 给定 $x\neq 0$，设$f(cx):=c\|x\|$，对所有的 $c\in K$，在由 x 生成的一维子空间 J 上定义 f. 那么 f 在 J 上具有所期望的性质. 由哈恩-巴拿赫定理，它可扩张到所有的 X 中，并保持 $\|f\|'=1$. 在$\|\cdot\|''$的定义中，利用这个 f 给出了 $\|I''x\|''\geqslant x$，所以 $\|I''x\|=x$. □

定义 一个赋范线性空间$(X,\ \|\cdot\|)$称为是**自反的**，当且仅当 I''取由 X 映上到 X'. [192]

在任意情况下，I''是一个线性等距同构(保持由范数定义的度量)且是 1－1 的. 因此，每个有限维的赋范空间是自反的. 因为对偶空间(如 X'')是完备的(由定理6.1.3)，所以一个自反的赋范空间一定是巴拿赫空间. 由定理 5.5.1 可推出希尔伯特空间都是自反的.

这里有一个非自反空间的例子. 回顾 ℓ^1 是所有实数序列$\{x_n\}$的集合，使得范数

$$\|\{x_n\}\|_1:=\sum_n |x_n|<\infty.$$

因此，ℓ^1 是具有正整数集合上的计数测度的 ℓ^1. 令 y 是对偶空间$(\ell^1)'$的任意元素. 令 e_n 是第 n 个位置是 1，其余是 0 的序列. 设 $y_n:=y(e_n)$. e_n 的有限线性组合在 ℓ^1 中是稠密的，从 y 到$\{y_n\}$的映射是 1－1 的. 令 ℓ^∞ 是所有有界实数序列$\{y_n\}$的集合，且由上确界范数 $\|\{y_n\}\|_\infty:=\sup_n|y_n|$. 那么$\ell^1$ 的对偶和 ℓ^∞ 是一致的，且 $\|y\|_1'=\|\{y_n\}\|_\infty$.

令 c 为由所有收敛序列形成的 ℓ^∞ 的子空间. 在 c 上，线性型 f 定义为 $f(y):=\lim_{n\to\infty}y_n$，则 $\|f\|_\infty'\leqslant 1$ 且由哈恩-巴拿赫定理，f 可扩张为$(\ell^\infty)'$中的一个元素. 现在，f 不在 ℓ^1 上的 I''的值域中，故 ℓ^1 不是自反的.

习题

1. 下列哪些函数在$\mathbb{R}$ 的指定子集上是利普希茨的？
   (a)$f(x)=x^2$，$x\in[0,\ 1]$.
   (b)$f(x)=x^2$，$x\in[1,\ \infty)$.
   (c)$f(x)=x^{1/2}$，$x\in[0,\ 1]$.
   (d)$f(x)=x^{1/2}$，$x\in[1,\ \infty)$.
2. 设 f 是 X 上的一个利普希茨函数且 E 是 X 的子集，其中 X 上的 $\|f\|_L$ 与限制到 E 上的是相同的. 设 $x\in X\setminus E$，E 上的 f 唯一地决定 X 上的 f 吗？
   (a)证明：如果 x 在 E 的闭包 $\bar{E}$ 中，那么对此问题的回答是肯定的.
   (b)给出对某个 f，即使 x 不在 $\bar{E}$ 中，回答也是肯定的例子.
   (c)举一个使得回答是否定的 X，E，f 和 x 的例子.
3. 证明：一个巴拿赫空间 X 是自反的，当且仅当它的对偶 X'是自反的. 注："必要性"容易证明. [193]
4. 设$(X,\ \|\cdot\|)$是一个线性赋范空间，E 是一个线性闭子空间，且 $x\in X\setminus E$. 证明：对某个 $f\in X'$，在 E 上 $f\equiv 0$，且 $f(x)=1$.
5. 证明：利普希茨函数的扩张定理(6.1.1)对复值函数不一定成立. [提示：设 $S=\mathbb{R}^2$，范数为 $\|(x,\ y)\|:=$

$\max(|x|, |y|)$. 令 $E:=\{(0,0),(0,2),(2,0)\}$且$f(0,0)=0$, $f(2,0)=2$, $f(0,2)=1+3^{1/2}\mathrm{i}$. 怎样将$f$扩张到$(1,1)$?]

6. 设(S,d)是一个度量空间, $E\subset S$且f是定义在E上一个复值利普希茨函数, 对E中所有的x和y, $|f(x)-f(y)|\leqslant Kd(x,y)$. 证明: f可以扩张到所有的S中, 且对S中所有的u和v, 有$|f(u)-f(v)|\leqslant 2^{1/2}Kd(u,v)$.
7. 令c_0为所有实数序列$\{x_n\}$形成的空间, 且当$n\to\infty$时, $\{x_n\}$收敛到0, 范数为$\|\{x_n\}\|_s:=\sup_n|x_n|$. 证明: $(c_0,\|\cdot\|_s)$是一个巴拿赫空间, 它的对偶空间和ℓ^1空间是等距的, ℓ^1是所有可和序列形成的空间(L^1是具有计数测度的整数空间). 证明: c_0不是自反的.
8. 设X是一个巴拿赫空间且对偶空间为X', T是X的一个子集. 令$T^{\perp}:=\{f\in X': f(x)=0, \forall t\in T\}$. 同样, 对$S\subset X'$, 令$S^{\perp}:=\{x\in X: f(x)=0, \forall f\in S\}$. 证明: 对任意的$T\subset X$,
(a)$T^{\perp}$是X'的一个线性闭子空间.
(b)$T\subset(T^{\perp})^{\perp}$.
(c)$T=(T^{\perp})^{\perp}$, 如果T是一个线性闭子空间. [提示: 见习题4.]
9. 设$(X,\|\cdot\|)$是一个巴拿赫空间且对偶空间为X', 则X'上的弱*拓扑(weak * topology)是最小的拓扑, 使得对每个$x\in X$, 由$I''(x)(f):=f(x)$ $(f\in X')$定义的函数I''在X'上是连续的. 证明: $E'_1:=\{y\in X': \|y\|'\leqslant 1\}$关于弱*拓扑(Alaoglu定理)是紧的. [提示: 利用吉洪诺夫定理并证明在X上的所有实函数形成的集合中, E'_1关于内积拓扑是闭的.]
10. 设$(X,\mathcal{S},\mu)$是一个测度空间, $(S,\|\cdot\|)$是一个可分的巴拿赫空间. 从X映射到S的函数f称为是简单的, 如果它在$\mathcal{S}$中的每个集合上只取有限个值. 设g是从X映射到S的任意可测函数, 满足$\int\|g\|\mathrm{d}\mu<\infty$. 证明: 存在简单函数$f_n$, 满足当$n\to\infty$时, $\int\|f_n-g\|\mathrm{d}\mu\to 0$. 如果$f$是简单的且$f(x)=\sum_{1\leqslant i\leqslant n}1_{A(i)}(x)s_i$, $s_i\in S$且$A(i)\in\mathcal{S}$, 令$\int f\mathrm{d}\mu=\sum_{1\leqslant i\leqslant n}\mu(A(i))s_i\in S$. 证明: $\int f\mathrm{d}\mu$是明确定义的且如果$\int\|f_n-g\|\mathrm{d}\mu\to 0$, 则$\int f_n\mathrm{d}\mu$收敛到$S$中的一个仅依赖于$g$的元素, 记为$\int g\mathrm{d}\mu$(Bochner积分).

194

11. (连续性). 如果g从X映射到S的函数, 则t称为g的Pettis积分, 如果对所有的$u\in S'$, $\int u(g)\mathrm{d}\mu=u(t)$. 证明: g的Bochner积分(如在习题10中所定义的)当它存在时, 它也是Pettis积分.
12. 证明: 复数域$\mathbb{C}$上的每个希尔伯特空间H是自反的. [提示: 令$C(h)(f):=(f,h)$, $f,h\in H$. 那么由定理5.5.1, C取H映上到H'. 用(5.3.3)证明对所有的$h\in H$, $\|C(h)\|'=\|h\|$. 对所有的f, $h\in H$, 令$(C(f),C(h))':=(h,f)$. 证明定义$H'\times H'$上的一个内积$(\cdot,\cdot)'$, 使得对所有的$\psi\in H'$, $[(\psi,\psi)']^{1/2}=\|\psi\|'$, 则$H'$就是一个希尔伯特空间. 在$H'$及$(H')'$上应用定理5.5.1得, I''是$(H')'$上的简单函数.]

6.2 凸集及其分离性

令V是一个向量空间. X中的一个子集是凸的, 当且仅当对任意的x, $y\in C$, $\lambda\in[0,1]$, 有$\lambda x+(1-\lambda)y\in C$. 那么$\lambda x+(1-\lambda)y$称为$x$和$y$的一个凸组合(convex combination). 同样, C称为是一个锥(cone), 如果对所有的$x\in C$, 对所有的$t\geqslant 0$, 有$tx\in C$. (如图6-2所示). 对向量空间X上的任意半范数$\|\cdot\|$, $r\geqslant 0$和$y\in X$, 集合$\{x: \|x\|\leqslant r\}$和$\{x: \|x-y\|\leqslant r\}$都是凸的.

称V中的集合A在$x\in A$是径向的(radial), 当且仅当对每个$y\in V$, 有一个$\delta>0$, 当$|t|<\delta$时, 有$x+ty\in A$. 因此, 在通过x的每条直线L上, $A\cap L$包含了L中的一个开区间, 对一条直线的

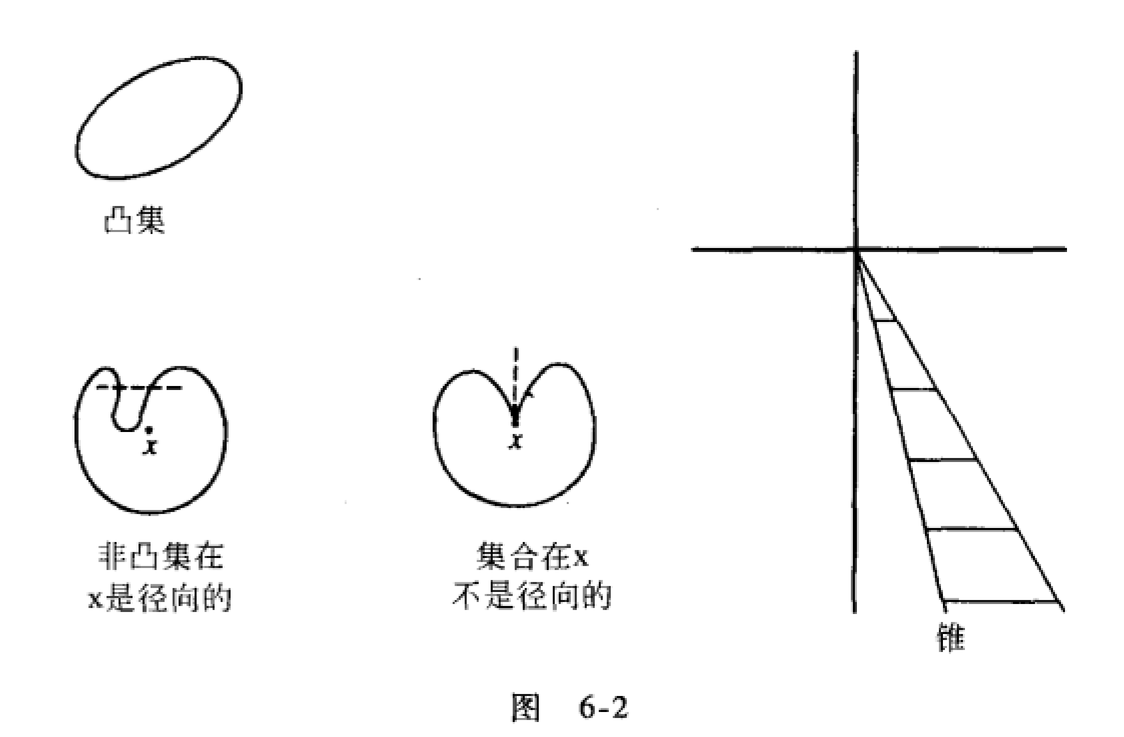


图 6-2

通常拓扑而言，它包含了 x. 集合 A 称为是径向的，当且仅当它在它的每个点是径向的. 通常把关于径向的结论应用于某个拓扑的开集，但它对完善一些与拓扑无关的结论也是很有用的.

例如，在 $\mathbb{R}^2$ 的极坐标 $(r,\ \theta)$ 中，令 A 是所有直线 $\{(r,\ \theta):\ r\geqslant 0\}$ 和线段 L_θ 的并，其中 θ/π 是无理数，这里如果 $0\leqslant\theta<2\pi$，$\theta/\pi=m/n$ 是有理数，且在最低项中，则 $L_\theta=\{(r,\ \theta):\ 0\leqslant r<1/n\}$. 这个集合 A 在 0 点是径向的，在其他处不是. [195]

一个集合 A 的径向核(radial kernel)定义为包含 A 的最大径向集 A_0. 显然，径向集的并是径向集，所以 A_0 是明确定义的，尽管它可能是空的. 注意到，A_0 不一定是集合 A_r，A_r 是使得 A 是径向的点的集合. 在刚给定的例子中，$A_r=\{0\}$，但 $\{0\}$ 不是一个径向集. 对于凸集，结果会更好一些.

6.2.1 定理 如果 A 是凸集，则 $A_r=A_0$.

证明 显然，$A_0\subset A_r$. 反之，设 $x\in A_r$，取 V 中任意的 w 和 z，则对某个 $\delta>0$，当 $|s|<\delta$ 时，$x+sw\in A$，且当 $|t|<\delta$ 时，$x+tz\in A$. 则由凸性 $(x+sw+x+tz)z\in A$ 知，当 $|a|<\delta/2$，$|b|<\delta/2$ 时，$x+aw+bz\in A$. 由于 z 是任意的，当 $|a|<\delta/2$ 时，A 在 $x+aw$ 是径向的，因此 $x+aw\in A_r$. 因为 w 是任意的，A_r 在 x 是径向的，所以 A_r 是一个径向集，$A_r\subset A_0$，从而 $A_r=A_0$. □

在 $\mathbb{R}^k$ 中，显然，任意集合的内部包含在它的径向核中. 通常径向核等于其内部. 一个例外是 $\mathbb{R}^2$ 中的集合 A，它的补由所有满足 $x>0$ 且 $x^2\leqslant y\leqslant 2x^2$ 的 $(x,\ y)$ 组成. 于是 $A_0=A$ 但是 A 不是开的，因为它不包含 $(0,\ 0)$ 的邻域. 如果 A 是凸集，则下面的引理表明了 A_0 是凸的并且可在 A 中任取 y.

6.2.2 引理 如果 A 是凸的且在 x 是径向的，则当 $y\in A$ 且 $0\leqslant t<1$ 时，$(1-t)x+ty\in A_0$.

证明 对每个 $z\in V$，取 $\varepsilon>0$，当 $|u|<\varepsilon$ 时，$x+uz\in A$. 令 $v:=(1-t)u$，则由凸性，

$$(1-t)(x+uz)+ty=((1-t)x+ty)+vz\in A.$$

这里，v 可能是满足 $|v|<(1-t)\varepsilon$ 的任意数. 因此，$(1-t)x+ty$ 在 A_r 中，由引理 6.2.1，A_r 等于 A_0. □

下面的定理是本节的主要结论. 这个定理蕴涵着两个不相交的凸集，其中至少有一个在某处是径向的，可以分离成两个半空间，这个半空间由一个线性型 f 定义，除了有界超平面 $\{x:\ f(x)=c\}$ 外，$\{x:\ f(x)\geqslant c\}$ 和 $\{x:\ f(x)\leqslant c\}$ 是不相交的. [196]

6.2.3 分离定理 设 A 和 B 是一个实向量空间 V 的非空凸子集，使得 A 在某个点 x 是径向的且

$A_0 \cap B = \varnothing$，则存在 V 上的一个实线性函数 f，不恒为 0，使得

$$\inf_{b \in B} f(b) \geq \sup_{a \in A} f(a).$$

注：由引理 6.2.2，A_0 是非空的．因为 $f \neq 0$，存在某个 $f(z) \neq 0$ 的 z．那么随着 t 的变化，$f(x+tz) = f(x) + tf(z)$，所以 f 不是 A 上的常数．这里称 f 分离 A 和 B．例如，令 $A = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leqslant 1\}$，$B = \{(1, y): y \in \mathbb{R}\}$，则 $A_0 = \{(x, y): x^2 + y^2 < 1\}$ 和 B 是不相交的，且一个分离线性泛函 f 是由 $f(x, y) = cx$ 给定的，其中 $\forall c > 0$．

证明 令 $C := \{t(b-a): t \geqslant 0, b \in B, a \in A\}$，则 C 是一个凸锥．由引理 6.2.1，$x \in A_0$．取 $y \in B$．如果 $x - y \in C$，取 $u \in A$，$v \in B$ 和 $t \geqslant 0$，使得 $x - y = t(v-u)$，则 $x + tu = y + tv$ 且 $(x+tu)/(1+t) = (y+tv)/(1+t)$．由引理 6.2.2，$(x+tu)/(1+t) \in A_0$，但 $(y+v)/(1+t) \in B$，得到矛盾．所以 $x - y \notin C$．因此，$C \neq V$．令 $z := y - x$，故 $-z \notin C$．引理 6.2.4 和引理 6.2.5 是定理 6.2.3 证明的一部分．

6.2.4 引理 对任意的向量空间 V，凸锥 $C \subset V$ 和 $p \in V \setminus C$，有一个包含 C 但不包含 p 的最大凸锥 M．对每个 $v \in V$，$v \in M$ 或 $-v \in M$（或 $v, -v \in M$）．

例：设 $V = \mathbb{R}^2$，$C = \{(\xi, \eta): |\eta| \leqslant \xi\}$ 且 $p = (-1, 0)$，则 M 是一个半平面 $\{(\xi, \eta): \xi \geqslant c\eta\}$，其中 $|c| \leqslant 1$．

证明 如果 C_α 是一个凸锥，由包含，它是线性的，但不包含 p，则它们的并也是一个不包含 p 的凸锥．所以，由佐恩引理，有一个包含 C 不包含 p 的最大凸锥．对任意的凸锥 M 及 $v \in V$，$\{tv + u: u \in M, t \geqslant 0\}$ 是包含 M 和 v 的最小凸锥．因此，如果后一个结论在引理 6.2.4 中不成立，对某个 $u \in M$，$t > 0$，有 $p = u + tv$．同样，对某个 $w \in M$，$s > 0$，有 $p = w - sv$，则

$$(s+t)p = s(u+tv) + t(w-sv) = su + tw = (s+t)\left(\frac{su}{s+t} + \frac{tw}{s+t}\right) \in M,$$

因为 M 是一个凸锥．但因为 $1/(s+t) > 0$，所以 $p = (s+t)p/(s+t) \in M$，得出矛盾，故引理 6.2.4 成立． □

197

对在定理 6.2.3 证明开始所定义的 C 和 z，$(z = y - x)$，集合 $\{y - u: u \in A\}$ 在 z 是径向的，故由引理 6.2.1 知，$z \in C_0$．令 $p := -z$．由引理 6.2.4，取一个最大的凸锥 $M \supset C$ 且 $p \notin M$，则 $z \in M_0$．

称向量空间 V 的一个线性子空间 E 有余维数（codimension）k，当且仅当有一个 k-维线性子空间 F，使得 $V = E + F := \{\xi + \eta: \xi \in E, \eta \in F\}$，$k$ 是使这个式子成立的最小维数．

对引理 6.2.4 的例子，$M \setminus M_0$ 是直线 $\xi = c\eta$，此直线是一个余维数为 1 的线性子空间．这使得更一般的结论成立．

6.2.5 引理 如果 $z \in M_0$，其中 M 是一个不包含 $-z$ 的最大的凸锥，则 $M \setminus M_0$ 是 V 的一个余维数为 1 的线性子空间．

证明 对任意的 $u \notin M$．对某个 $a > 0$ 及 $m \in M$，$au + m = -z$，那么 $-au/2 = (z+m)/2$，所以由引理 6.2.2 知，M 在 $-au/2$ 是径向的且 $t = 1/2$．因为 M 是一个锥，所以 M 在 $-u$ 是径向的．故由引理 6.2.1，$-u \in M_0$．如果 $0 \in M_0$，则 $M = V$，和 $-z \notin M$ 矛盾．因此 $0 \notin M_0$．对任意的 $v \in M_0$，我们得到 $-v \notin M$，否则，由引理 6.2.2 知，$0 = (v + (-v))/2 \in M_0$．因此，$M_0 = \{-u: u \notin M\}$．

如果 $v \in M \setminus M_0$，则 $-v \in M \setminus M_0$．于是 $M \cap -M = M \setminus M_0$．既然 M 是一个凸锥，$M \cap -M$ 是一个线性子空间．为了证明它有余维数 1，对 1 维的子空间 $\{cz: c \in \mathbb{R}\}$，任取 $g \in V$．如果 $g \in -M_0$，由于 M_0 及 $-M_0$ 是径向的，则 $S := \{t: tg + (1-t)z \in M_0\}$ 和 $T := \{t: tg + (1-t)z \in -M_0\}$ 是非空的实

数开集且 $0\in S$，$1\in T$. 由于 S 和 T 是不相交的，它们的并不包含所有的 $[0,1]$ 及一个连通集 ($S\cap[0,1]$ 的上确界不包含在 S 或 T 中). 因此，对某个 t，$0<t<1$，$tg+(1-t)z\in M\cap -M$. 于是正如所期望的对某个 $w\in M\cap -M$，$g=-(1-t)z/t+w$. 另一方面，如果 $g\notin -M_0$，则 $g\in M$，且 $g\in M\cap -M=M\setminus M_0$ 或 $g\in M_0$，故 $-g\in -M_0$ 且 $-g$ 在 $M\cap -M$ 和 $\{z\}$ 生成的线性空间中，因此 g 也在 $M\cap -M$ 和 $\{z\}$ 生成的线性空间中. 故对所有的 g，此结论都成立，从而证明了 $M\cap -M$ 有余维数 1，因此引理 6.2.5 成立. □

现在来定义 V 上的一个线性型 f，它满足在 $M\cap -M$ 上 $f(z)=1$ 及 $f=0$. 那么，在 $M\cap -M$ 上 $f=0$，且在 M_0 上 $f>0$，但在 $-M_0$ 上，$f<0$. 则对所有的，$b\in B$ 和 $a\in A$，$f(b)\geqslant f(a)$，这就证明了定理 6.2.3. □

注：*在定理的其他假设下，A_0 和 B 不相交这个条件对一个分离函数 f 的存在性也是必要的，因为如果 $x\in A_0\cap B$ 且 $f(v)>0$，则对足够小的 c，$x+cv\in A$ 且 $f(x+cv)>f(x)$.*

198

给定一个拓扑空间 X 中的一个集合 S，其闭包为 $\bar{S}$，内部为 int S，回顾 S 的边界定义为 $\partial S:=\bar{S}\setminus \mathrm{int}S$. 在一个线性赋范空间 X 中，一个闭的超平面(hyperplane)定义为集合 $f^{-1}\{c\}$，其中 f 是某个连续的线性型，不恒为 0 且 c 是常数. 对一个集合 $S\subset X$，一个在点 $x\in\partial S$ 的支撑超平面(support hyperplane)是一个包含 x 的闭的超平面 $f^{-1}\{c\}$，使得在 S 上，或者 $f\geqslant c$ 或者 $f\leqslant c$. 例如，在 $\mathbb{R}^2$ 中，集合 S(其中 $y\leqslant x^2$)没有支撑超平面，在这种情况下，其支撑超平面将是一条直线，因为通过边界上一点的每条直线在它的两边上都有 S 的点. 另一方面，边界的切线是 S 的余的一个支撑超平面.

注意到在 $\mathbb{R}^k$ 上，所有的线性型是连续的，故超平面是闭的. 下面将讨论关于 $\mathbb{R}^k$ 中的凸集合的某些结论.

6.2.6 定理 *对于 $\mathbb{R}^k$ 中任意的凸集合 C，要么 C 有非空的内部，要么 C 包含在某个超平面中.*

证明 任取 $x\in C$，令 W 是所有的 $y-x(y\in C)$ 生成的线性空间(包含 $y-x$ 的最小的线性子空间). 如果 W 不全在 $\mathbb{R}^k$ 中，则有一个在 $\mathbb{R}^k$ 上不恒为 0 但在 W 上为 0 的线性函数 f，那么 C 包含在超平面 $f^{-1}\{f(x)\}$ 中.

反之，如果 $W=\mathbb{R}^k$，设 y_j-x 是线性无关的，其中 $j=1,\cdots k$ 且 $y_j\in C$. 所有的凸组合 $p_0x+\sum_{1\leqslant j\leqslant k}p_jy_j$(其中 $p_j\geqslant 0$ 且 $\sum_{0\leqslant j\leqslant k}p_j=1$ 的集合称为是一个单纯形(simplex). (例如，如果 $k=2$，它是一个矩形.)那么 C 包含这个单纯形，且有非空的内部，这个内部由那些对所有的 j，$p_j>0$ 的点组成. □

如果一个凸集合是一个岛，则从岸上的每个点，至少有海洋的一个 180°的无障碍视角. 下面把这个结论扩张到一般的凸集合.

6.2.7 定理 *对 $\mathbb{R}^k$ 中任意的凸集合 C 和任意的 $x\in\partial C$，C 在 x 至少有一个支撑超平面 $f^{-1}\{c\}$. 如果 y 在 C 的内部且在 C 上 $f\leqslant c$，则 $f(y)<c$.*

注：*由支撑超平面的定义，在 C 上或者 $f\leqslant c$ 或者 $f\geqslant c$. 如果必要，用 $-f$ 代替 f，$-c$ 代替 c，我们可假定在 C 上 $f\leqslant c$.*

证明 令 $x\in\partial C$. 如果 C 有空的内部，则由定理 6.2.6，C 就包含在某个超平面中，从而 C 就是一个支撑超平面. 故假设 C 有包含点 y 的非空内部，则 C 在 y 是径向的. 令 $g(t):=x+t(x-y)$.

199

如果对某个 $t>0$，有 $g(t)\in C$，注意到 $x=\dfrac{ty}{1+t}+\dfrac{g(t)}{1+t}$ 是 y 和 $g(t)$ 的一个凸组合. 如果在 C 中，y 可由 y 的邻域中的每个点代替，由于 $t>0$，则同样的凸组合就给出了 x 的一个邻域中所有的点，且此

凸组合包含在 C 中，与 $x\in\partial C$ 矛盾. 因此，如果我们令 $B:=\{g(t): t>0\}$，则 $B\cap C=\varnothing$ 且 $B\cap C_0=\varnothing$.

因此，在集合 $C(C_0\neq\varnothing)$ 和 B 中应用分离定理(6.2.3)，得到一个非零线性型 f 且 $\inf_B f\geqslant\sup_C f$. 因 $x\in\bar{C}$，f 是一致连续的，$f(x)\leqslant\sup_C f$，并且 $x\in B$ 推得 $f(x)\geqslant\inf_B f$. 因此 $\inf_B f=f(x)=\sup_C f$. 令 $c:=f(x)$ 且令 H 是超平面 $\{u\in\mathbb{R}^k: f(u)=f(x)\}$，则 H 是 C 在 x 处的支撑超平面. 如果 $f(y)\geqslant C$，则由于 f 不是常数，在 y 的每个邻域中它可能有比 c 大的值，因此在 C 上，得到矛盾. 故 $f(y)<c$. □

6.2.8 命题 *对于上述证明中的 f 和 g，对所有的 $t>0$，$f(g(t))>f(x)$.*

证明 由于 $f(y)<f(x)$，则可推得 $f(g(t))$ 是 t 的严格递增函数，从而命题成立. □

$\mathbb{R}^k$ 中的一个闭半空间(half-space)定义为集合 $\{x: f(x)\leqslant c\}$，其中 f 是一个非零线性型且 $c\in\mathbb{R}$. 注意到 $\{x: f(x)\geqslant c\}=\{x: -f(x)\leqslant -c\}$，也是一个闭半空间. 容易看出任意的半空间是凸的，故半空间的任意交是凸的. 反之，这里有 $\mathbb{R}^k$ 中的闭凸集合的一个性质.

6.2.9 定理 *$\mathbb{R}^k$ 中的集合 A 是闭的且凸的，当且仅当 A 是一个闭半空间的交.*

例：$\mathbb{R}^2$ 中的凸 k 边形是 k 个半空间的交. 圆盘 $\{(x, y): x^2+y^2\leqslant 1\}$ 是所有半空间 $\{(x, y): sx+ty\leqslant 1\}$ 的交，其中 $s^2+t^2=1$.

证明 显然，闭半空间的交是闭的且凸的. 相反，令 C 是闭的且凸的. 如果 $k=1$，则 C 是一个闭区间(两个半空间的交)或半直线或整条直线. 整条直线是半空间的空集的交(由定义)，故对 $k=1$ 结论成立. 对 k 作归纳. 如果 C 无内部，则由定理 6.2.6，它包含在一个超平面 $H=f^{-1}\{c\}$ 中，对某个非零线性型 f 和 $c\in\mathbb{R}$，则

200

$$H=\{x: f(x)\geqslant c\}\cap\{x: f(x)\leqslant c\}$$

是两个半空间的交. 同样，对某个 $(k-1)$ 维的线性子空间 V 和 $u\in\mathbb{R}^k$，

$$H=u+V:=\{u+v: v\in V\}.$$

由归纳假设，$C-u$ 是半空间 $\{v\in V: f_\alpha(v)\leqslant c_\alpha\}$ 的一个交. V 上的线性型 f_α 可以假定在 $\mathbb{R}^k$ 上有定义. 于是

$$C=\{x\in H: \text{对所有的 }\alpha,\ f_\alpha(x)\leqslant c_\alpha+f_\alpha(u)\},$$

故 C 是闭半空间的一个交.

因此，假设 C 有内部，故 C_0 非空，可说包含一个点 y. 对不属于 C 的每个点 z，连接 y 和 z 的线段 L 一定交 ∂C 于某个点 x，因为 C 的内部和余是开集，且每个开集与 L 有非空的交(如引理 6.2.5 的证明). 由定理 6.2.7，设 $f^{-1}\{c\}$ 是 C 在 x 的一个支撑超平面，这里我们可以取 C 包含在 $\{u: f(u)\leqslant c\}$. 那么由命题 6.2.8，对某个 $t>0$，$z=g(t)$，$f(z)>f(x)=c$，$z\notin\{u: f(u)\leqslant c\}$. 故所有这样的半空间 $\{u: f(u)\leqslant c\}$ 的交恰好是 C.

$\mathbb{R}$ 中的两个相邻但不相交的开区间的并，例如，$(0, 1)\cup(1, 2)$ 不是凸的且比它的闭包的内部小. 对于凸集，后一种情况不会出现.

***6.2.10 命题** *$\mathbb{R}^k$ 中任意的凸开集 C 是其闭包 $\bar{C}$ 的内部.*

证明 每个开集都包含在它的闭包的内部中. 假设 $x\in(\operatorname{int}\bar{C})\setminus C$. 对 C 和 $\{x\}$ 应用分离定理(6.2.3)，所以有一个非零线性型 f，使得 $f(x)>\sup_{y\in C}f(y)$. 由于 $\operatorname{int}\bar{C}$ 是开的，则在 $\operatorname{int}\bar{C}$ 中存在某个 u，且 $f(u)>f(x)$. 由于 f 是连续的，有 $v_n\in C$ 且 $f(v_n)\to f(u)$，故 $f(v_v)>f(x)$，得出矛盾. □

接下来，有一个介于分离定理(6.2.3)和哈恩－巴拿赫定理(6.1.4)之间的一个结论，就是下面的习题 6.

201

***6.2.11 定理** 设 X 是一个线性空间，E 是一个线性子空间. 设 U 是 X 的一个凸子集且在 E 中的某个点是径向的. 设 h 是 E 上的一个非零线性型且在 $U\cap E$ 上有上界. 那么 h 可以扩张为 X 上的一个实线性型 g 且 $\sup_{x\in U} g(x)=\sup_{y\in U\cap E} h(y)$.

证明 令 $t:=\sup_{y\in U\cap E} h(y)$ 和 $B:=\{y\in E: h(y)>t\}$，则 U 和 B 都是凸的且不相交. 对 $A=U$ 应用分离定理(6.2.3)，给出 X 上的一个非零线性型且 $\sup_{x\in U} f(x)\leqslant \inf_{v\in B} f(v)$.

设 U 在 $x_0\in U\cap E$ 是径向的，则 $f(x_0)<\sup_{x\in U} f(x)$，故对任意的 $v\in B$，$f(v)>f(x_0)$ 且 f 在 E 上不是常数(0). 令 $F:=\{x\in E: f(x)=f(x_0)\}$. 假设 h 在 F 中的点上取两个不同的值. 于是在连接这些点的直线上，h 取所有的实值，故直线与 B 相交. 但是在直线上，$f\equiv f(x_0)$，矛盾. 因此，h 在 F 上是常数，即在 F 上 $h\equiv c$. 包含 F 和 $E\setminus F$ 中任意点 w 的最小线性子空间是 E. 如果 $c=0$，那么由于 h 在 E 上不是常数，我们得到 $0\in F$ 且 $f(x_0)=0$. 取 $w\in E\setminus F$，对 $x=w$ 和所有的 $x\in F$，有 $h(w)\neq 0\neq f(w)$ 及 $f(x)=f(w)h(x)/h(w)$. 因此，对所有的 $x\in E$，也同样成立. 另一方面，如果 $c\neq 0$，则 0 不在 F 中，故对所有的 $x\in F$ 及 $x=0$，$f(x_0)\neq 0$ 且 $f(x)=f(x_0)h(x)/c$. 因此，对所有的 $x\in E$ 成立. 故在另一种情况下，对某个 $\alpha\neq 0$，在 E 上 $f\equiv\alpha h$. 由于 f 和 h 在 x_0 处的值比在 B 上的值要小，$\alpha>0$. 令 $g:=f/\alpha$，则 g 是一个线性型，使得 h 扩张到 X 上，且

$$\sup_{x\in U} g(x)\leqslant \inf_{v\in B} f(v)/\alpha=\inf_{v\in B} h(v)=t.$$

□

习题

1. 举例：$\mathbb{R}^2$ 的一个集合 A 在区间 $\{(x,0): |x|<1\}$ 上的每个点都是径向的，但使得这区间不包含在 A 的内部中.
2. 证明：椭球面 $x^2/a^2+y^2/b^2+z^2/c^2\leqslant 1$ 是凸的. 求它的边界的每个点处的一个支撑面并证明它是唯一的.
3. 在 $\mathbb{R}^3$ 中，通过原点的哪些平面是单位立方 $[0,1]^3$ 的支撑平面？求出此立方的所有支撑平面.
4. 所有收敛到 0 的实数序列的巴拿赫空间 c_0 有范数 $\|\{x_n\}\|:=\sup_n|x_n|$. 证明：在单位球面 $B:=\{y: \|y\|\leqslant 1\}$ 的边界 ∂B 的一点 x 处的每个支撑超平面包含 ∂B 的其他点. [提示：见 6.1 节的习题 7.]
5. 举例：一个巴拿赫空间 X，X 中的一个闭凸集 C，一个点 $u\in X$ 且在 C 中无唯一的邻近点. [提示：令 $X=\mathbb{R}^2$ 且范数为 $\|(x,y)\|:=\max(|x|,|y|)$.]

202

6. 设 $(X,\|\cdot\|)$ 是一个实赋范空间，E 是一个线性子空间且 $h\in E'$. 给出证明：h 可扩张为以定理 6.2.11 为基础的 X' 中的一个元素(哈恩–巴拿赫定理 6.1.4). [提示：令 $U:=\{x\in X: \|x\|\leqslant 1\}$. (由于这样的关系，对凸集的分离定理有时称为哈恩－巴拿赫定理的“几何形式”.)]
7. 证明：在任意的有限维巴拿赫空间(有任意范数的 $\mathbb{R}^k$)，对任意的闭凸集 C 和不在 C 中的任意点 x，至少在 C 中有一个邻近的点 y，换句话说，$\|y-x\|=\inf_{z\in C}\|z-x\|$.
8. 举例：巴拿赫空间 S 中的一个闭凸集 C 和一个点 $x\in S$ 且此点在 C 中无邻近的点. [提示：令 S 是具有绝对可和序列的 ℓ^1 空间且范数为 $\|\{x_n\}\|_1:=\sum_n|x_n|$. 令 $C:=\left\{\{t_j(1+j)/j\}_{j\geqslant 1}: t_j\geqslant 0, \sum_j t_j=1\right\}$ 且 $x=0$.]
9. (“几何数”). 向量空间 V 中的一个子集称为是**对称的**，当且仅当 $X\in C$ 时有 $-x\in C$. 假设 C 是 $\mathbb{R}^k$ 中的一个凸的对称集且勒贝格体积为 $\lambda(C)>2^k$. 证明：C 至少包含一个点 $z=(z_1,\cdots,z_k)\in\mathbb{Z}^k$，即 z 有整数坐标 $z_i\in\mathbb{Z}$，对所有的 i，且 $z\neq 0$. [提示：每个 $x\in\mathbb{R}^k$ 可写作 $x=y+z$，其中 $z/2\in\mathbb{Z}^k$ 且对所有的 i，$-1<y_i\leqslant 1$. 证明：C 中一定有 x 且 $x'\neq x$，且对同样的 y，考虑 $(x-x')/2$.]
10. 一个开半空间是由 $\{x: f(x)>c\}$ 这样的集合构成的，其中 f 是一个连续线性函数. 证明：在 $\mathbb{R}^k$ 中，任意的开凸集是开半空间的交.

6.3 凸函数

设 V 是一个实向量空间，C 是 V 中的一个凸集. C 上的一个实值函数 f 称为凸的，当且仅当对 C 中的每个 x 和 y 以及 $0 \leqslant \lambda \leqslant 1$，

6.3.1
$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y).$$

所有有序对 $\langle \lambda x + (1-\lambda)y,\ \lambda f(x) + (1-\lambda)f(y)\rangle$ $(0 \leqslant \lambda \leqslant 1)$ 称为是 f 图形上连接两个点 $\langle x, f(x)\rangle$ 和 $\langle y, f(y)\rangle$ 的一条弦(chord). 因此，一个凸函数 f 是一条在 f 图形上或其上方的弦. 例如，在 $\mathbb{R}$ 上 $f(x) = x^2$ 是一个凸函数，但 $g(x) = -x^2$ 不是凸的.

在 $\mathbb{R}$ 中，一个凸集是一个区间(可能是闭的或是开的，在每个端点有界的或无界的). 凸函数有如下“递增的微商”.

203

6.3.2 命题 *设 f 是 $\mathbb{R}$ 中区间 J 上的一个凸函数，在 J 中 $t < u < v$. 那么*
$$\frac{f(u)-f(t)}{u-t} \leqslant \frac{f(v)-f(t)}{v-t} \leqslant \frac{f(v)-f(u)}{v-u}.$$

证明 (如图 6-3A 所示)令 $x := t$ 和 $y := v$，则我们发现 $u = \lambda x + (1-\lambda)y$，对 $\lambda = (v-u)/(v-t)$，其中 $0 < \lambda < 1$，应用(6.3.1)给出
$$f(u) \leqslant \frac{v-u}{v-t}f(t) + \frac{u-t}{v-t}f(v).$$

t u v

图 6-3A

如上所述的斜率之间的关系在图中是显然的. 更多的分析：后一个不等式可写为
$$(v-t)f(u) \leqslant (v-u)f(t) + (u-t)f(v),$$
从而
$$(v-t)(f(u)-f(t)) \leqslant (u-t)(f(v)-f(t)),$$
这给出了命题 6.3.2 中左边的不等式. 右边的不等式就是
$$(v-u)(f(v)-f(t)) \leqslant (v-t)(f(v)-f(u)),$$
再消去 $vf(v)$ 也成立. □

凸函数不一定是处处可微的. 例如，$f(x) := |x|$ 在 0 处不可微，尽管它有左右导数. 除了在其定义域的端点外，凸函数总有有限的一阶导数.

6.3.3 推论 *对定义在包含 $[a, b]$ 且 $a < b$ 的区间上的任意凸函数 f，随着 $h \downarrow 0$，右微商 $(f(a+h) - f(a))/h$ 是非递增的，有极限*

204

$$f'(a^+) := \lim_{h\downarrow 0}(f(a+h) - f(a))/h \geqslant -\infty.$$

如果 f 定义在任意 $t < a$ 上，则所有上述微商，和其极限至少是 $(f(a)-f(t))/(a-t)$，使得 $f'(a^+)$ 是有限的. 否则，$h \downarrow 0$ 时，微商 $(f(b)-f(b-h))/h$ 是非递减的，有一个极限 $f'(b^-) \leqslant +\infty$，且如果 f 定义在某个 $c > b$ 处，有 $f'(b^-) \leqslant (f(c)-f(b))/(c-b)$.

进一步，在 $[a, b)$ 上，$f'(x^+)$ 是 x 的非递增函数，在 $(a, b]$ 上，$f'(x^-)$ 是非递减的且对 $[a, b]$ 中所有的 x，有 $f'(x^-) \leqslant f'(x^+)$.

在 (a, b) 上，由于左右导数都存在且是有限的，故 f 是连续的.

证明 这些性质由命题 6.3.2 可直接得到. □

例： 设 $f(0) := f(1) := 1$ 及 $f(x) := 0$，$0 < x < 1$. 那么 f 是 $[0, 1]$ 上的凸函数但在端点不连续. 同样，对所有的 t 及 $x = 0$，令 $f(t) := t^2$. 那么右导数是 $h^2/h = h$，当 $h \downarrow 0$，它趋于 0. 当 $h \downarrow 0$ 时，左

导数 $-h^2/h=-h\uparrow 0$.

$\mathbb{R}^k$ 上的一个限制取值在一条直线上的凸函数 f 是凸的，所以由推论 6.3.3，一定有一阶方向导数. 这些导数将被证明其绝对值是有界的，在 f 有定义的一个开集合的紧子集上是一致有界的.

6.3.4 定理 *设 f 是 $\mathbb{R}^k$ 中开凸集 U 上的一个实值函数. 那么在 U 的每个点 x 及每个 $v\in\mathbb{R}^k$ 处，f 在方向 v 处有一个有限的方向导数，*

$$D_v f(x) := \lim_{h\downarrow 0}(f(x+hv)-f(x))/h.$$

在包含 U 的任意紧凸集 K 上，且 v 是有界的，如 $|v|\leqslant 1$，这些方向导数是有界的，且 f 在 K 上是利普希茨的. 因此，f 在 U 上是连续的.

证明 直接由推论 6.3.3，方向导数是存在的. 剩下的证明对 k 用归纳法. 对 $k=1$，设 $a<b<c<d$ 且在 (a,d) 上 f 是凸的. 设 $a<t<b$ 及 $c<v<d$. 那么 $[b,c]$ 上的所有 f 的导数的下界是 $(f(b)-f(t))/(b-t)$，上界是 $(f(v)-f(c))/(v-c)$，由命题 6.3.2，所以在 $[b,c]$ 上 f 是利普希茨的且由推论 6.3.3，左右导数在 $[b,c]$ 上是一致有界的.

205

在高维的情形中，假设 C 是一个包含在 U 中的闭立方体 $\prod_{j=1}^{k}[c_j,c_j+s]$. 那么对某个 $\delta>0$，C 包含在另一个包含在 U 中的闭立方体 $D:=\prod_{j=1}^{k}[c_j-\delta,c_j+s+\delta]$ 中. 另外，这些立方体在 f 是凸函数的开集的内部是低维立方体，f 是利普希茨的，因此由归纳假设，f 是连续有界的. 因此，对某个 $M<\infty$，$\sup_{\partial D}f-\inf_{\partial C}f\leqslant M$ 且 $\sup_{\partial C}f-\inf_{\partial D}f\leqslant M$. 对 int C 中任意的两个点 r 和 s，设直线 L 通过 r 和 s 与 ∂D 交于 p 点，然后和 ∂C 依次交于 q，r，s，t，再交 ∂D 于 u（如图 6-3B 所示）. 则

$$-\frac{M}{\delta}\leqslant\frac{f(q)-f(p)}{|q-p|}\leqslant\frac{f(s)-f(r)}{|s-r|}\leqslant\frac{f(u)-f(t)}{|u-t|}\leqslant\frac{M}{\delta}.$$

（为看出第二个不等式，例如，可以插入一个中间项 $(f(r)-f(q))/|r-q|$，再应用命题 6.3.2.）因此，$|f(s)-f(r)|\leqslant M|s-r|/\delta$ 且 f 在 C 上是利普希茨的. 故在 int$C$ 上对 $|v|\leqslant 1$，有 $|D_v f|\leqslant M/\delta$.

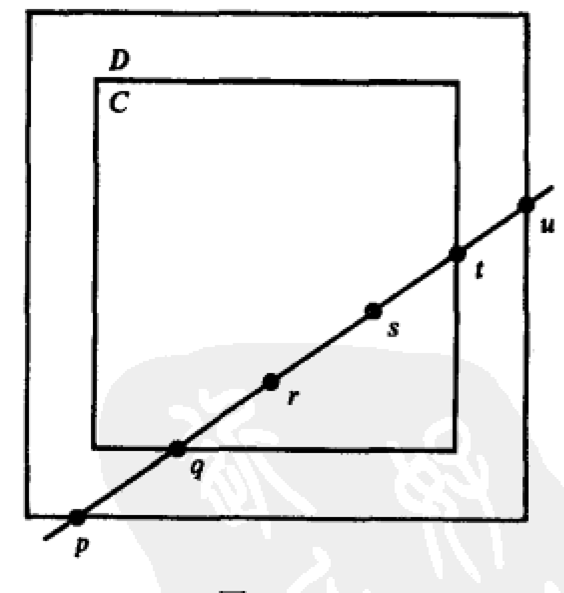


图 6-3B

现在，任意包含在 U 中的紧凸集 K 有通过这样的立方体 C 的内部的一个开覆盖，且有一个通过立方体 $C_i(i=1,\cdots,n)$ 内部的一个有限子覆盖. 取有限多个边界的最大值，有一个 $N<\infty$，使得对 $|v|=1$，方向导数 $D_v f$ 在某个 C_i 上以 N 为长度中是有界的，因此在 K 上也是有界的. 于是得到 f 在 K 上是利普希茨的，且对所有的 $x,y\in K$，$|f(x)-f(y)|\leqslant N|x-y|$. 对每个 $x\in U$ 及某个 $t>0$，$|y-x|<t$ 蕴涵着 $y\in U$. 然后取 $K:=\{y:|y-x|\leqslant t/2\}$ 是紧的且包含在 U 中，f 在 x 是连续的，故 f 在 U 上是连续的.

例：设在 $(0,\infty)$ 上 $f(x):=1/x$，则 f 是凸的且连续的，并在闭子集 $[c,\infty)(c>0)$ 上是利普希茨的，但 f 在 $(0,\infty)$ 上不是利普希茨的且为了在 0 连续，f 不能定义.

206

在 10.2 节中关于凸函数有进一步的结论.

习题

1. 假设 $\mathbb{R}$ 中开区间 J 上的一个实值函数 f 在 J 上有二阶导数 f''. 证明：f 是凸的当且仅当在 J 上处处有 $f''\geqslant 0$.

2. 设对所有的 x 和 y，$f(x,\ y):=x^2y^2$. 证明：虽然对每个 y，f 在 x 是凸的，对每个 x，f 在 y 是凸的，但在 $\mathbb{R}^2$ 上 f 不是凸的.
3. 如果 f 和 g 是同一个定义域中的两个凸函数，证明：$f+g$ 和 $\max(f,\ g)$ 都是凸的. 举例说明 $\min(f,\ g)$ 不一定是凸的.
4. 对任意集合 F 和一个度量空间中的点 x，回顾 $d(x,\ F):=\inf\{d(x,\ y):\ y\in F\}$. 令 F 是一个具有通常的距离 $d(x,\ y):=\|x-y\|$ 的线性赋范空间 S 中的一个闭集. 证明：$d(\cdot,\ F)$ 是一个凸函数当且仅当 F 是一个凸集合.
5. 设 U 是 $\mathbb{R}^k$ 中的一个凸开集. 对任意集合 $A\subset\mathbb{R}^k$，令 $-A:=\{-x:\ x\in A\}$. 假设 $U=-U$，所以 $0\in U$. 设 μ 是定义在 U 的博雷尔子集上的一个测度且 $0<\mu(U)<\infty$，$\mu(B)=\mu(-B)$. 设 f 是 U 上的一个凸函数. 证明：$f(0)\leqslant\int f\mathrm{d}f\mu/\mu(U)$. [提示：用像测度定理 4.2.8 且使 $T(x)\equiv -x$.]
6. 设 f 是 $\mathbb{R}^2$ 中一个凸开集上的实函数，存在二阶偏导数 $D_{ij}f:=\partial^2 f/\partial x_i\partial x_j$ 且对 i 及 $j=1,\ 2$，它在 U 中是处处连续的. 证明：如果矩阵 $\{D_{ij}f\}_{i,j=1,2}$ 在 U 中的每个点上都是非负定的，f 是凸的. [提示：考虑 f 到直线的限制，再用习题 1 的结果.]
7. 证明：对定义在一个包含闭区间 $[a,\ b]$ 的开区间上的任意凸函数 f, $f(b)-f(a)=\int_a^b f'(x^+)\mathrm{d}x=\int_a^b f'(x^-)\mathrm{d}x$.
8. 假设在凸函数的定义中允许取值 $-\infty$. 设 f 是定义在凸集 $A\subset\mathbb{R}^k$ 上的一个凸函数且对某个 $x\in A$，$f(x)=-\infty$. 证明：对所有的 $y\in A$，$f(t)=-\infty$.
9. 设 $f(x):=|x|^p$，对所有的实数 x. p 取什么值时，下列结论成立？
(a) f 是凸的；(b) 对所有的 x，微分 $f'(x)$ 存在；(c) 对所有的 x，二阶导数 $f''(x)$ 存在.
10. 设 f 是定义在凸集 $A\supset\mathbb{R}^k$ 上的一个凸函数. 对 $\mathbb{R}^k$ 中某个固定的 x 和 y，令 $g(t):=f(x+ty)+f(x-ty)$，不论何时都是有定义的. 证明：g 是定义在 $\mathbb{R}$ 中一个开区间(可能为空)上的非递减函数.

207

11. 回顾在 $\mathbb{R}$ 上的一个*拉东测度*是一个从 μ 映射到 $\mathbb{R}$ 的函数 μ，定义在所有的有界博雷尔集上且在任意固定的有界区间的博雷尔子集上是可数可加的. 证明：对定义在开区间 $U\subset\mathbb{R}$ 上的任意凸函数 f，存在唯一的一个非负拉东测度 μ，使得对所有 U 中的 $x<y$，$f'(y^+)-f'(x^+)=\mu((x,\ y])$ 和 $f'(y^-)-f'(x^-)=\mu([x,\ y))$.
12. (连续性). 相反地，证明：对开区间 $U\subset\mathbb{R}$ 上任意的非负拉东测度 μ，存在 U 上的一个凸函数 f，满足习题 11 中的关系. 如果 g 是另一个这样的凸函数(对同一个 μ)，$f-g$ 必须满足什么样的条件？

*6.4 L^p 空间的对偶性

回顾从一个希尔伯特空间 H 映射到它的标量域($\mathbb{R}$ 或 $\mathbb{C}$)的任意连续线性函数 F，可以写为一个内积：对 H 中的某个 g，对 H 中所有的 h，$F(h)=(h,\ g)$(定理 5.5.1). 特别地，如果对某个测度 μ，H 是一个 $L^2(X,\ \mathcal{B},\ \mu)$ 空间，那么对 $\mathcal{L}^2$ 中的某个 g，对 L^2 中所有的 h，L^2 上的每个连续线性型 F 可写为 $F(h)=\int h\,\bar{g}\mathrm{d}\mu$. 相反，每个这样的积分定义了一个连续线性型，由 Cauchy-Bunyakovsky-Schwarz 不等式(5.3.3)，$\left|\int(h_1-h_2)\,\bar{g}\mathrm{d}\mu\right|\leqslant\|h_1-h_2\|\ \|g\|$. 由 Rogers-Höder 不等式(5.1.2)，以同样的方法，这些结论可扩展到 $L^p(p\neq 2)$ 空间上. 也就是说如果 $1/p+1/q=1$，$p\geqslant 1$，而且如果 $f\in\mathcal{L}^p(\mu)$，$g\in\mathcal{L}^q(\mu)$，那么 $\left|\int fg\mathrm{d}\mu\right|\leqslant\|f\|_p\|g\|_q$. 所以令 $F(f)=\int fg\mathrm{d}\mu$，$g\in\mathcal{L}^q(\mu)$，在 L^p 上定

义了一个连续线性型 F，因为 $\left|\int(f_1-f_2)g\mathrm{d}\mu\right|\leqslant\|f_1-f_2\|_p\|g\|_q$. 对 $p=q=2$，我们注意到每个连续线性型都可以由这种方式表示. 这种表示可扩展到 $1\leqslant p<\infty$（一般对 $p=\infty$ 不成立）. 下一个定理，则给出了逆 Rogers-Höder 不等式.（回顾定义在 6.1 节中的对偶巴拿赫空间和对偶范数 $\|\cdot\|$ 的概念）这个定理给出对偶空间 $(L^p)'$ 和 L^q 的一个等距同构. 在这种意义下，对 $1\leqslant p<\infty$，L^p 的对偶是 L^q.

6.4.1 Riesz 表示定理 *对任意 σ-有限测度空间 $(X,\mathcal{S},\mu)$，$1\leqslant p<\infty$，且 $p^{-1}+q^{-1}=1$，由*

$$T(g)(f):=\int fg\mathrm{d}\mu$$

定义的映射 T 给出了一个 $(L^q,\|\cdot\|_q)$ 映上到 $((L^p),\|\cdot\|_p)$ 的等距同构.

208

证明 首先假设 $1<p<\infty$，则由 Rogers-Höder 不等式(5.1.2)，取 T 为 L^q 到 $(L^p)'$ 的一个映射，且 $\|T(g)\|_p'\leqslant\|g\|_q$. 给定 $\mathcal{L}^q$ 中的 g，如果 $g(x)\neq 0$，令 $f(x):=|g(x)|^q/g(x)$，如果 $g(x)=0$，令 $f(x):=0$. 那么，$\int fg\mathrm{d}\mu=f|g|^q\mathrm{d}\mu$ 及 $\int|f|^p\mathrm{d}\mu=\int|g|^{p(q-1)}\mathrm{d}\mu=\int|g|^q\mathrm{d}\mu$，所以 $\|T(g)\|_p'\geqslant\left(\int|g|^q\mathrm{d}\mu\right)^{1-1/p}=\|g\|_q$. 所以对 $\mathcal{L}^q$ 中所有的 g，$\|T(g)\|_p'=\|g\|_q$，T 是其中的等距同构，即对 $\mathcal{L}^q$ 中所有的 g 和 γ，$\|T(g)-T(\gamma)\|_p'=\|g-\gamma\|_q$.

现在假设 $L\in(L^p)'$. 如果 $K=\mathbb{C}$，则对实 $(L^p)'$ 中的某个 M 和 N，对实 $\mathcal{L}^p$ 中所有的 f，有 $L(f)=M(f)+\mathrm{i}N(f)$. 故为了证明 T 是 $(L^p)'$ 上的等距同构，我们假设 $K=\mathbb{R}$.

对 $f\geqslant 0$，$f\in L^p$，令 $L^+(f):=\sup\{L(h):0\leqslant h\leqslant f\}$. 因为 $0\leqslant h\leqslant f$ 蕴涵了 $\|h\|_p\leqslant\|f\|_p$，$L^+(f)$ 是有限的. 显然对任意的 $c\geqslant 0$，有 $L^+(cf)=cL^+(f)$.

6.4.2 引理 *如果 $g\geqslant 0$，$f_1\geqslant 0$ 及 $f_2\geqslant 0$ 且这三个函数都在 $\mathcal{L}^p$ 中，那么 $0\leqslant g\leqslant f_1+f_2$ 当且仅当对某个测度 g_i 且 $0\leqslant g_i\leqslant f_i$，$i=1, 2$，$g=g_1+g_2$.*

证明 “充分性”显然. 为证“必要性”，假设 $0\leqslant g\leqslant f_1+f_2$. 令 $g_1:=\min(f_1, g)$ 及 $g_2:=g-g_1$，则显然对 $i=1, 2$，$g_i\geqslant 0$ 且 $g_1\leqslant f_1$. 现在，$g_2\leqslant f_2$，因为否则在某个 x 处，$g_2\geqslant f_2\geqslant 0$ 就推出了 $g_1=f_1$，故 $f_1+f_2<g_1+g_2=g$，得到矛盾，从而证明了引理. □

显然，如果对 $i=1, 2$，$f_i\geqslant 0$ 则 $L^+(f_1+f_2)\geqslant L^+(f_1)+L^+(f_2)$. 反之，由引理推出 $L^+(f_1+f_2)\leqslant L^+(f_1)+L^+(f_2)$，所以 $L^+(f_1+f_2)=L^+(f_1)+L^+(f_2)$. 我们定义 $L^+(f_1-f_2)=L^+(f_1)-L^+(f_2)$. 这是明确定义的，因为如果 $f_1-f_2=g-h$，其中 g 和 h 也是 $\mathcal{L}^p$ 中的非负函数，那么 $f_1+h=f_2+g$，故 $L^+(f_1)+L^+(h)=L^+(f_1+h)=L^+(f_2+g)=L^+(f_2)+L^+(g)$，且 $L^+(f_1)-L^+(f_2)=L^+(g)-L^+(h)$.

因此，L^+ 在 L^p 上是有定义的且是线性的. 令 $L^-:=L^+-L$，则 L^- 是线性的，在 $\mathcal{L}^p$ 中对所有的 $f\geqslant 0$，$L^-(f)\geqslant 0$ 且 $L=L^+-L^-$.

$\mathcal{L}^p$ 是一个斯通向量格（如 4.5 节中定义的）. 如果对所有的 x，$f_n(x)\downarrow 0$ 且 $f_1\in\mathcal{L}^p$，那么由控制或单调定理(4.3 节)，用 $p<\infty$，$\|f_n\|_p\to 0$. 所以由斯通–丹尼尔定理(4.5.2)，有测度 β^+ 及 β^-，且对所有的 $f\in\mathcal{L}^p$，$L^+(f)=\int f\mathrm{d}\beta^+$ 及 $L^-(f)=\int f\mathrm{d}\beta^-$. 显然，$\beta^+$ 及 β^- 都关于 μ 是绝对连续的. 首先假设 μ 是有限的. 那么 β^+ 及 β^- 也是有限的（令 $f=1$），故由 Radon-Nikodym 定理(5.5.4)，有非负可测函数 g^+ 及 g^- 且 $\beta^+(A)=\int_A g^+\mathrm{d}\mu$，对 β^-，g^- 及任意可测集合 A 有类似的结论. 视简单函数和

209

它们的单调极限为通常形式(命题 4.1.5)，我们得到对所有的$f\in\mathcal{L}^p$，$L^+(f)=\int fg^+\mathrm{d}\mu$(首先考虑$f\geqslant 0$，则通常令$f=f^+-f^-$). 同样对所有的$f\in\mathcal{L}^p$，$L^-(f)=\int fg^-\mathrm{d}\mu$. 设$g:=g^+-g^-$. 函数$g$可以一致地定义在有限$\mu$-测度集合序列上，且这些序列的并(几乎)全在$X$中，尽管$g$的可积性目前仅在(一些)有限测度集上成立. 当在某些有限测度集之外$f=0$时，我们得到$L(f)=\int fg\mathrm{d}\mu$.

为了证明$g\in\mathcal{L}^q$，对$|g|\leqslant n$，令$g_n:=g(n):=g$，其他情况下，令$g_n:=0$. 对集合$E_n:=E(n)$且$\mu(E)<\infty$，令$f_n:=1_{E(n)}1_{g(n)\neq 0}|g_n|^q/g_n$. 那么$f_n\in L^p$且

$$\|L\|'_p\geqslant |L(f_n)|/\|f_n\|_p=\left(\int_{E(n)}|g_n|^q\mathrm{d}\mu\right)^{1-1/p}=\|1_{E(n)g_n}\|_q.$$

令$n\to\infty$，因为μ是σ-有限的，我们得到$E(n)\uparrow X$. 因此，由法图引理(4.3.3)，$g\in\mathcal{L}^q$. 那么由控制收敛定理(4.3.5)及 Rogers-Hölder 不等式(5.1.2)知，对所有的$f\in\mathcal{L}^p$，$L(f)=\int fg\mathrm{d}\mu$. 这就证明了$1<p<\infty$的情况.

剩下的情况就是$p=1$，$q=\infty$. 如果$g\in\mathcal{L}^\infty$，则$T(g)$在$(L^1)'$中且$\|T(g)\|'_1\leqslant\|g\|_\infty$. 对任意的$\varepsilon>0$，有一个集合$A$且$0<\mu(A)<\infty$，对所有的$x\in A$，$|g(x)|>\|g\|_\infty-\varepsilon$. 我们可假设$\varepsilon<\|g\|_\infty$(如果$\|g\|_\infty=0$，则$g=0$a.e.). 令$f:=1_A|g|/g$，则$f\in\mathcal{L}^1$且$\left|\int fg\mathrm{d}\mu\right|\geqslant(\|g\|_\infty-\varepsilon)\|f\|_1$. 令$\varepsilon\downarrow 0$，推出$\|T(g)\|'_1\geqslant\|g\|_\infty$，因此对$p>1$，$T$是由$\mathcal{L}^\infty$映射到$(L^1)'$的一个等距同构.

为了证明T是映上的，设$L\in(L^1)'$. 和以前一样，我们可以定义分解$L=L^+-L$且找到了一个可测函数g，使得对$\mathcal{L}^1$中所有的在一些有限测度集外是 0 的那些f，有$L(f)=\int fg\mathrm{d}\mu$. 如果$g$不在$\mathcal{L}^\infty$中，则对所有的$n=1,2,\cdots$有一个集合$A_n:=A(n)$，满足$0<\mu(A_n)<\infty$且对$A_n$中所有的$x$，$|g(x)|>n$. 令$f:=1_{A(n)}|g|/g$. 那么对所有的$n$，

$$\|L\|'_1\geqslant|L(f)|/\|f\|_1=\left|\int fg\mathrm{d}\mu\right|\Big/\mu(A_n)\geqslant n,$$

得到矛盾. 所以$g\in\mathcal{L}^\infty$且T是映上的. □

习题

1. (a)证明：$L^1([0,1],\lambda)$不是自反的. [提示：$C[0,1]\subset\mathcal{L}^\infty\subset(L^1)'$；如果$T(f):=f(0)$，则$T\in C[0,1]'$. 用哈恩-巴拿赫定理(6.1.4).]

[210] (b)证明：$\ell^1:=L^1(\mathbb{N},2^{\mathbb{N}},c)$不是自反的，其中$c$是$\mathbb{N}$上的计数测度.

2. 回顾 5.2 节习题 1 中定义的空间$L^p(X,\mathcal{S},M)$，$0<p<1$.
   (a)证明：对$X=[0,1]$，$\mu=\lambda$，只有一个从$\mathcal{L}^p$到$\mathbb{R}$的连续线性函数恒为 0.
   (b)X是无穷集且μ为计数测度，证明：对$0<p<1$，在$\ell^p:=L^p(X,2^X,\mu)$上存在非零连续线性实值函数.
3. 设S是集合X上的实值函数的一个向量格，如在 4.5 节中定义的. 设L是一个由S到$\mathbb{R}$的线性函数. 在S中对$f\geqslant 0$，令$L^+(f):=\sup\{L(g):0\leqslant g\leqslant f\}$.
   (a)证明：对某个$f\in S$，$L^+(f)$可能是无穷的. [提示：参见命题 5.6.4.]
   (b)如果$L^+(f)$是有限的，对所有的$f\geqslant 0$，证明：L^+及$L^-:=L^+-L$可扩张为从S到$\mathbb{R}$的线性函数.
4. 对$1\leqslant r\leqslant p<\infty$，设$V$由$L^p$映射到$L^r$是恒等的，其中对$s=p$或$r$，$L^s:=L^s([0,1],\lambda)$. 设$U$是$V$从$(L^r)'$

映射到$(L^p)'$的一个转置，对每个$h\in(L^r)'$，$U(h):=h\circ V$. 证明：U不是映上到$(L^p)'$的函数.［提示：考虑函数$h(x):=1/x^{\alpha}$，$0<x\leqslant 1$，$0<\alpha<1$.］

6.5 一致有界性及闭图形

本节将要证明经典泛函分析中的四个重要定理中的三个（第四个是哈恩-巴拿赫定理6.1.4）. 设$(X,\ \|\cdot\|)$和$(Y,\ |\cdot|)$是两个线性赋范空间. 一个从X映射到Y的线性函数T通常称为一个线性变换(transformation)或算子. 如果X和Y是有限维的，且X有一组线性无关基$e_1,\ \cdots,\ e_m$，Y有一组基$f_1,\ \cdots,\ f_n$，则T就确定了，且由矩阵$\{T_{ij}\}_{1\leqslant i\leqslant n,1\leqslant j\leqslant m}$确定，其中$T(e_j)=\sum_{1\leqslant i\leqslant n}T_{ij}f_i$. 对无穷维的巴拿赫空间，本节的这个事实是有意义的，其中即使某类的基可能存在（例如，希尔伯特空间的规范正交基，5.3节），根据其基和矩阵，算子通常研究的很少. 回顾（由定理6.1.2）：一个线性算子T是连续的，当且仅当它在

$$\|T\|:=\sup\{|T(x)|:\|x\|\leqslant 1\}<\infty$$

的意义下是有界的. 这里$\|T\|$称为T的算子范数(operator norm). 第一个重要的定理是，巴拿赫空间上的连续线性算子的逐点有界集在算子范数之下也有界的. 对一个有限维空间来说，这是不足为奇的，在这个有限维空间上可取一组有限基的最大值. 一般地，对于无穷维空间它有更多的优越性. 211

6.5.1 定理（一致有界性原理） *设$(X,\ |\cdot|)$是一个巴拿赫空间. 对指标集I中的每个α，设T_α是从X映射到某个线性赋范空间$(S_\alpha,\ |\cdot|_\alpha)$的一个有界线性算子. 假设对$X$中的每个$x$，$\sup_{\alpha\in I}|T_\alpha(x)|_\alpha<\infty$. 那么$\sup_{\alpha\in I}\|T_\alpha\|<\infty$.*

证明 设$A:=\{$对所有的$\alpha\in I$，$x\in X$，$|T_\alpha(x)|_\alpha\leqslant 1\}$. 那么由假设，$\bigcup_n nA=X$，其中$nA:=\{nx:\ x\in A\}$. 由范畴定理(2.5.2)，某个$nA$在某些地方是稠密的. 因为乘$n$的乘法是同态的，所以$A$在某些地方是稠密的. 现在，$A$是闭的，所以对某个$x\in A$及$\delta>0$，球

$$B(x,\delta):=\{y:|y-x|<\delta\}$$

包含在A中. 因为A是对称的，$B(-x,\ \delta)$也包含在A中. 同时A是凸的，所以对任意的$u\in B(0,\ \delta)$，

$$u=\frac{1}{2}(x+u+(-x+u))\in A.$$

这就得出对所有的α，$\|T_\alpha\|\leqslant 1/\delta$. □

例：设H是一个实的希尔伯特空间且有规范正交基$\{e_n\}_{n=1}^{\infty}$. 设X是所有e_n的有限线性组合形成的集合，是一个不完备的内积空间. 对所有的x，设$T_n(x)=(x,\ ne_n)$，则对所有的$x\in X$，当$n\to\infty$时，$T_n(x)\to 0$，且$\sup_n|T_n(x)|<\infty$. 对所有的n，设$S_n=\mathbb{R}$，则当$n\to\infty$时，$\|T_n\|=n\to\infty$. 这表明完备性的假设在定理6.5.1中是如此的重要.

在定理6.5.1的大多数应用中，空间S_α都是相同的. 通常它们都等于域K，以使得$T_\alpha\in X'$. 在那种情况下，定理即是说对任意的巴拿赫空间X，如果对X中任意的x，有$\sup_\alpha|T_\alpha(x)|<\infty$，则$\sup_\alpha\|T_\alpha\|'<\infty$.

F-空间(F-space)是域$K(=\mathbb{R}$或$\mathbb{C})$上的一个向量空间V，且有一完备的度量d，d是不变的：对V中所有的$x,\ y,\ z$，$d(x,\ y)=d(x+z,\ y+z)$，对V中的加法及K中的标量乘法对d而言是联合连续的. 显然，一个具有通常度量的巴拿赫空间是一个F-空间.

从一个拓扑空间映射到另一个拓扑空间的函数 T 称为开的，当且仅当对它的定义域中的每个开集，$T(U):=\{T(x): x\in U\}$ 在值域空间中是开的. 在下面的定理中，假设 T 映上到 Y 的算子(不只映射到 Y)和 Y 是完备的，这个假设是至关重要的. 例如，设 H 是希尔伯特空间且有规范正交基 212 $\{e_n\}_{\geq 1}$. 设 T 是一个线性算子且对所有的 n，$T(e_n)=e_n/n$，则 T 是连续的且 $\|T\|=1$，但不是映上到 H 的函数(当然映上到它的值域，它是不完备的)且不是开的.

6.5.2 开映射定理 设(X, d)和(Y, e)都是 F-空间. 设 T 是一个从 X 映上到 Y 的连续线性函数. 那么 T 是开的.

证明 对 $r>0$，令 $X_r:=\{x\in X: d(0, x)<r\}$，同样地，令 $Y_r:=\{y\in Y: e(0, y)<r\}$. 对一个给定的固定值 r，令 $V:=X_{r/2}$. 那么，由标量乘法的连续性，$\bigcup_n nV=X$. 所以由范筹定理(2.5.2)，某个集合 $T(nV)=nT(V)$ 在 Y 中的某个非空开集中稠密，且 $T(V)$ 在某个球 $y+Y_\varepsilon(\varepsilon>0)$ 中稠密. 现在 V 是对称的，因对任意的 x，$d(0, x)=d(-x, 0)=d(0, -x)$，且 $V-V:=\{u-v: u\in V, v\in V\}$ 包含在 X_r 中. 因此 $T(X_r)$ 在 Y_ε 中稠密.

现在设 s 是比 r 大的任意数. 设 $r_1:=r$，$r_n:=(s-r)/2^{n-1}$，$n\geq 2$. 那么 $\sum_n r_n=s$. 对每个 j，令 $r(j):=r_j$. 于是对每个 $j=1, 2, \cdots$，对某个 $\varepsilon_j:=\varepsilon(j)>0$，$T(X_{r(j)})$ 在 $Y_{\varepsilon(j)}$ 中稠密. 下面将证明 $T(X_s)$ 包含 Y_ε. 给定任意的 $y\in Y_\varepsilon$，取 $x_1\in X_{r(1)}$，使得 $e(T(x_1), y)<\varepsilon_2$. 然后选择 $x_2\in X_{r(2)}$，使得 $e(T(x_2), y-T(x_1))<\varepsilon_3$，所以 $e(T(x_1+x_2), y)<\varepsilon_3$. 因此，递归地，对 $j=1, 2, \cdots$ 存在 $x_j\in X_{r(j)}$，使得对 $n=1, 2, \cdots$，$e(T(x_1+\cdots+x_n), y)<\varepsilon_{n+1}$.

由于 X 是完备的，$\sum_{j=1}^{\infty} x_j$ 在 X 中收敛到某个 x 且 $T(x)=y$，$x\in X_s$，故 $T(X_s)\supset Y_\varepsilon$.

因此，对 X 中 0 的任意邻域 U 及 $x\in X$，$T(U)$ 包含 Y 中 0 的一个邻域，且 $T(x+U)$ 包含 $T(x)$ 的一个邻域. 所以 T 是开的. □

接下来的定理比开映射定理应用更广泛，从开映射定理更容易得出.

6.5.3 闭图形定理 设 X 和 Y 是 F-空间，T 是一个从 X 映射到 Y 的一个线性算子. 那么 T 是连续的当且仅当它(即对 $X\times Y$ 上的乘积拓扑，它的图形 $\{\langle x, Tx\rangle; x\in X\}$)是闭的.

证明 "必要性"显然，因为如果 $\langle x_n, T(x_n)\rangle\to\langle x, y\rangle$，那么由连续性，$T(x_n)\to T(x)=y$. 相反，假设 T 是闭的，则作为 F-空间 $X\times Y$ 的一个线性闭子空间，它是自身的一个 F-空间. 映上射影 $P(x, y):=x$ 把 T 映上到 X 且是连续的. 所以由开映射定理(6.5.2)，P 是开的. 既然它是 213 1-1 的，它就是一个同胚. 因此，它的逆变换 $x\to\langle x, T(x)\rangle$ 是连续的，故 T 是连续的. □

6.5.4 推论 对任意从一个 F-空间 X 映上到另一个 F-空间 Y 的连续线性 1-1 的算子 T，其逆 T^{-1} 也是连续的.

证明 这是开映射定理(6.5.2)或闭图形定理的一个推论($Y\times X$ 中 T^{-1} 的图形是闭的，在 $X\times Y$ 中 T 的图形也是闭的). □

再者，假设 T 是映上到值域空间 Y，且假设 Y 是完备的，这两个假设是很重要的，见下面的习题 2.

习题

1. 证明：对任意两个线性赋范空间$(X, \|\cdot\|)$及$(Y, |\cdot|)$，设 $L(X, Y)$ 是所有从 X 映射到 Y 的有界线性算子的集合，算子范数 $T\mapsto\|T\|$ 事实上是 $L(X, Y)$ 上的一个范数.

2. 详细证明：满足 $T(e_n)=e_n/n$ 的算子 T 是连续的但不是开的，不是映上的，其中 $\{e_n\}$ 是希尔伯特空间的一组规范正交基. 证明它的值域是稠密的.
3. 证明：在开映射定理(6.5.2)中，只需假设 T 的值域是 Y 中的第二范筹而不需要 T 是到 Y 上的满射，其中第二范筹是指不是有限个处处稠密子集的并.
4. 设$(X, \|\cdot\|)$及$(Y, |\cdot|)$是巴拿赫空间. 设 T_n 是从 X 映射到 Y 的有界线性算子序列，使得对 X 中所有的 x，$T_n(x)$收敛到某个 $T(x)$. 证明：T 是一个有界线性算子.
5. 证明：习题 4 中的 T 不需要是有界的，如果序列 T_n 由 $\{T_\alpha\}_{\alpha\in I}$代替. 事实上，证明从 X 映射到 Y 的每个线性变换(连续或不连续)是某个有界线性算子网格的极限. [提示：用佐恩引理，证明 X 是一个哈梅尔基 B. 即 X 中的每个点是 B 中元素的唯一的有限线性组合. 用哈梅尔基证明有从任意个无限维的巴拿赫空间映射到$\mathbb{R}$ 上的无界线性函数.
6. 带有拓扑$\mathcal{T}$的一个实向量空间(线性空间)S 称为拓扑向量空间(topological vector space)，当且仅当加法从 $S\times S$ 映射到 S 是连续的，其标量乘法从$\mathbb{R}\times S$ 映射到 S 是连续的. 证明：$\mathbb{R}$ 上使其成为一个拓扑向量空间的仅有的豪斯多夫拓扑是通常的拓扑$\mathcal{U}$. [提示：对$\mathcal{U}$，在$\mathbb{R}$ 的紧子集上，用定理 2.2.11. 那么，如果$\mathcal{T}\neq\mathcal{U}$，对$\mathcal{T}$，有一个网格 $x_\alpha\to 0$，且 $|x_\alpha|\to\infty$，故 $1=(1/x_\alpha)x_\alpha\to 0$. [214]
7. 设 T^2 是所有复数序对(z, w)的集合，且有 $|z| = |w| = 1$. 设 f 是从$\mathbb{R}$ 映射到 T^2 的一个函数，f 定义为 $f(x)=(e^{ix}, e^{i\gamma x})$，其中 γ 是无理数，即 $\gamma=2^{1/2}$. 证明：f 是 1－1 的且是连续的但不是开的. 令$\mathcal{T}$是所有 $f^{-1}(U)$的集合，其中 U 是 T^2 中的开集. 证明：$\mathcal{T}$是一个拓扑，且对$\mathcal{T}$，其加法是连续的，但由$(\mathbb{R}, \mathcal{T})$到$(\mathbb{R}, \mathcal{T})$的标量乘法 $x\mapsto cx$ 不是连续的，甚至对一个给定的 c(如 $c=\sqrt{3}$). [提示：$\{e^{ix}, e^{i\gamma x}, e^{icx}: x\in\mathbb{R}\}$ 在 T^3 中稠密.]
8. 证明：$\mathbb{R}^k$ 上使其成为一个拓扑向量空间(对有限的 k)的仅有的豪斯多夫拓扑是通常的拓扑. [提示：$k=1$ 的情况就是习题 6.]
9. 设 F 是 F-空间 X 的一个有限维的线性子空间，所以存在某个有限集 $x_1, \cdots, x_n$，使得对某些数 a_j，每个 $x\in F$ 可表示为 $x=a_1x_1+\cdots+a_nx_n$. 证明：F 是闭的. [提示：用习题 8 和一致性的概念(2.7 节).]
10. 在任意无穷维 F-空间中，证明：任意的紧集有空的内部.
11. 一个从巴拿赫空间 X 映射到另一个巴拿赫空间 Y 的连续线性算子 T 称为紧的，当且仅当 $T\{x\in X: \|x\|\leqslant 1\}$ 有紧的闭包. 证明：如果 X 和 Y 是无穷维的，则 T 不是映上到 Y 的连续线性算子. [提示：应用习题 10.]
12. 证明：有一个可分的巴拿赫空间 X 和一个从 X 映射到自身的紧的线性算子 T，使得 $T\{x\in X: \|x\|\leqslant 1\}$ 不是紧的. [提示：令 $X=c_0$，所有收敛到 0 的序列形成的空间，且有范数 $\|\{x_n\}\|:=\sup_n|x_n|$.]
13. 设$(F_n, d_n)_{n\geqslant 1}$是 F-空间的任意序列. 证明：具有乘积拓扑的笛卡儿积 $\prod_n F_n$ (线性结构定义为$c\{x_n\}+\{y_n\}=\{cx_n+y_n\}$)是一个 F-空间. [提示：见定理 2.5.7.]
14. 在习题 13 中，如果每个 F_n 是具有通常的度量的$\mathbb{R}$，证明：在乘积空间上没有范数 $\|\cdot\|$，乘积空间上的度量为 $d(x, y):=\|x-y\|$.

*6.6 Brunn-Minkowski 不等式

对一个向量空间中的任意两集合 A 和 B，令 $A+B:=\{x+y: x\in A, y\in B\}$. 对任意的常数 c，令 $cA:=\{cx: x\in A\}$. 在$\mathbb{R}^k$ 中，如果 A 和 B 是紧的，则 $A+B$ 是紧的，是紧集 $A\times B$ 上的连续函数“＋”的值域. 因此，如果 A 和 B 是紧集的可数并，则 $A+B$ 也是紧集的一个可数并，故它是博雷尔可测的. 令 V 表示$\mathbb{R}^k$ 上的勒贝格测度(体积)，由定理 4.4.6 给出. 下面是关于凸性的一个重要定理. [215]

6.6.1 定理(Brunn-Minkowski 不等式)　设 A 和 B 是$\mathbb{R}^k$ 中的任意两个非空紧集，则

(a) $V(A+B)^{1/k} \geqslant V(A)^{1/k} + V(B)^{1/k}$.

(b) 对 $0<\lambda<1$，$V(\lambda A+(1-\lambda)B)^{1/k} \geqslant \lambda V(A)^{1/k} + (1-\lambda)V(B)^{1/k}$.

证明　显然，对 $c>0$，$V(cA)\equiv c^k V(A)$. 因此，(a)和(b)的特殊情况是等价的，其中 $\lambda=1/2$. 以 λA 代替 A，以 $(1-\lambda)B$ 代替 B，(b)可由(a)推出. 所以只需证明(a)，当 $V(A)=0$ 或 $V(B)=0$ 时，结论显然也成立.

因此，假设 A 和 B 有正的体积. $\mathbb{R}^k$ 中的两个集合 C 和 D 称为拟不相交(quasi-disjoint)，如果它们的交包含在某个$(k-1)$维的超平面 H 中. (例如，如果 $C\subset\{x: x_1\leqslant 3\}$，$D\subset\{x: x_1\geqslant 3\}$，则 C 和 D 是拟不相交的.)于是 $V(C\cup D)=V(C)+V(D)$. 这样的一个超平面 H 将 A 分为两个拟不相交闭集 A'，A''的并，它们是在 H 的每边上具有闭半空间的 A 的交. 同样地，设平行于 H 的超平面 J 将 B 分成两部分 B'，B''，选择使得 $r:=V(B')/V(B)=V(A')/V(A)$，其中 B'是 B 的交，B 是 J 的同一边上的闭半空间，类似地，A'是 A 的交，A 是 H 的同一边上的闭半空间. 在通常的$\mathbb{R}^k$ 坐标 $x=\langle x_1, \cdots, x_k\rangle$中，对 $c\in\mathbb{R}$，仅有超平面$\{x: x_i=c\}$，$i=1, \cdots, k$，这些在下面将用到. 如果 $H=\{x: x_i=c\}$，则对某个 d，$J=\{x: x_i=d\}$，且 $H+J=\{x: x_i=c+d\}$. 设 $A'=\{x\in A: x_i\leqslant c\}$，则有 $B'=\{x\in B: x_i\leqslant d\}$，等等.

现在，$A'+B'$和 $A''+B''$是拟不相交的，因 $A'+B'$包含在$\{x: x_i\leqslant c+d\}$中，$A''+B''$包含在$\{x: x_i\geqslant c+d\}$中，故

$$V(A)=V(A')+V(A''), V(B)=V(B')+V(B'') \quad \text{且}$$
$$V(A+B)\geqslant V(A'+B')+V(A''+B'').$$

令 $p(A, B):=V(A+B)-(V(A)^{1/k}+V(B)^{1/k})^k$. 因此我们只需证明 $p(A, B)\geqslant 0$. 我们有 $V(A'')/V(A)=1-r=V(B'')/V(B)$. 那么

$$s:=V(B)/V(A)=V(B')/V(A')=V(B'')/V(A'') \quad \text{且}$$
$$(V(A)^{1/k}+V(B)^{1/k})^k=V(A)(1+s^{1/k})^k.$$

对 A'，B'，A''，B''，相应的方程是成立的，则可推出

216

$$p(A,B)\geqslant p(A',B')+p(A'',B'').$$

一个分块(block)是闭区间的笛卡儿积. (在$\mathbb{R}^2$ 中，一个分块就是一个平行于坐标轴的闭矩形.)我们首先证明 $p(A, B)\geqslant 0$，其中 A 和 B 是分块. 设 A 的长度为 $a_1, \cdots, a_k$，B 的长度为 $b_1, \cdots, b_k$. 令$\alpha_i:=a_i/(a_i+b_i)$，$i=1, \cdots, k$ 及 $C:=A+B$. 那么 C 也是一个分块，且边长为 a_i+b_i，我们得到

$$p(A,B)=V(C)\{1-(\alpha_1\cdots\alpha_k)^{1/k}-((1-\alpha_1)\cdots(1-\alpha_k))^{1/k}\}.$$

因为几何平均小于算术平均(定理 5.1.6)，所以 $p(A, B)\geqslant 0$.

下面假设 A 是 m 个拟不相交分块的并，B 是 n 个这样分块的并. 对这种情况的证明可对 m 和 n 作归纳得到. 对 $m=n=1$ 已经证明了. 由归纳步，假设对至多 $m-1(m\geqslant 2)$个拟不相交分块的并以及至多是 n 时的并定理是成立的. 对任意两个拟不相交的分块，设为 C 和 D，在 A 的表示中，一定有某个 i，使得形成 C 的第 i 区间和 D 是拟不相交的，故对某个 c，C 或 D 一个包含在$\{x: x_i\leqslant c\}$中，且另一个包含在$\{x: x_i\geqslant c\}$中. 在具有这些半空间的 A 的表示中所有这些分块取交，我们得到了一个如上所述的分割 $A=A'\cup A''$，这里每个 A'和 A''是至多 $m-1$ 个不相交分块的并. 通过一个如上所描述的超平面 J，取 B 的一个相应的分割. 每个 B'和 B''仍然至多由 n 个分块组成. 故应用归纳

假设及

$$p(A,B) \geq p(A',B') + p(A'',B'') \geq 0,$$

从而完成了拟不相交分块的有限并的证明.

现在，任意的紧集 A 是一个拟不相交分块 U_n 的有限并的递减交集——例如，交于 A 的立方体的边界是区间 $[k_i/2^n, (k_i+1)/2^n]$，$k_i \in \mathbb{Z}$. 设 V_n 是相应于 B 的并. 那么 $U_n + V_n$ 递减到 $A+B$. 既然这些集合有有限的体积，应用单调收敛定理，收敛集合的体积收敛，从而给出任意紧集 A 和 B 在一般情况下的不等式.

Brunn-Minkowski 不等式还有另外的形式，可应用到凸集的截面. 设 C 是 $\mathbb{R}^{k+1}$ 中的一个集合，$u \in \mathbb{R}$，我们有 $\langle x, u\rangle \in \mathbb{R}^{k+1}$. 令 $C_u := \{x \in \mathbb{R}^k: \langle x, u\rangle \in C\}$. 设 V_k 表示 k-维的勒贝格测度.

6.6.2 定理（截面的 Brunn-Minkowski 不等式） *设 C 是 $\mathbb{R}^{k+1}$ 中的一个凸紧集. 令 $f(t) := V_k(C_t)^{1/k}$. 那么在 C 上 $f>0$ 的集合是一个区间 J，在此区间上 f 是凹的：对所有的 $u, v \in J$，且 $0 \leq \lambda \leq 1$，$f(\lambda u + (1-\lambda)v) \geq \lambda f(u) + (1-\lambda)f(v)$.*

217

证明 如果 $\langle x, u\rangle \in C$，$\langle y, v\rangle \in C$，则因 C 是凸的，故 $\langle \lambda x + (1-\lambda)y, \lambda u + (1-\lambda)v\rangle \in C$. 因此，

$$C_{\lambda u+(1-\lambda)v} \supset \lambda C_u + (1-\lambda)C_v.$$

然后由定理 6.6.1(b)就得到结论. □

习题

1. 证明：对某些闭集合 A 和 B，$A+B$ 不是闭的. [提示：令 $A = \{\langle x, y\rangle: xy=1\}$，$B = \{\langle x, y\rangle: xy=-1\}$.]
2. 设 C 是 $\mathbb{R}^{k+1}$ 中锥面的一部分，由
$$C := \{\langle x,u\rangle : x \in \mathbb{R}^k, u \in \mathbb{R}, |x| \leq u \leq 1\}$$
给出，对这样的 C，估计定理 6.6.2 中的函数 $f(t)$，并证明：$V_k(C_t)^p$ 是凹的当且仅当 $p \leq 1/k$，在 f 的定义中指数最有可能为 $1/k$.
3. 设 A 是 $\mathbb{R}^3$ 中顶点为 (0, 0, 0)，(1, 0, 0) 和 (1/2, 1, 0) 的三角形. 设 B 是一个顶点为 (1/2, 0, 1)，(0, 1, 1) 和 (1, 1, 1) 的三角形. 设 C 是 A 和 B 的凸包，对 $0 \leq t \leq 1$，估计 $V(C_t)$.

注释

6.1 节 哈恩-巴拿赫定理(6.1.4)是 Hahn(1927)和 Banach(1929, 1932)的著作中的结果. Banach(1932)对泛函分析做了很大贡献.

Lipschitz(1864)以他的名字定义了函数类. 关于利普希茨函数的扩张定理(6.1.1)的证明，本质上是哈恩-巴拿赫定理的最初证明中的一部分，但是非线性利普希茨函数的情况，好像最早是 Kirszbraun(1934)明确阐述的，由 McShane(1934)独立证明. Minty(1970)给出并修正了 Kirszbraun-McShane 定理的某些扩展.

6.2 节 凸集合(在二维或三维空间中)首先是由 H. Brunn(1887, 1889)系统研究的，然后是由 H. Minkowski(1897)提出. Brunn(1910)和 Minkowski(1910)证明了支撑(超)平面的存在性. Bonnesen 和 Fenchel(1934)，在哥本哈根，用大量的文献给出了更进一步的说明. Minkowski(1910)发现在数论中用凸集合，创造了"几何数".

凸集合的分离定理(6.2.3)应归功于 Dieudonné(1941). 更早，Eidelheit(1936)证明了在一个赋范线性空间中，"无普通内点"的凸集合的分离定理. 以上证明是建立在 Kelley-Namioka(1963, p. 19—23)

218 基础之上. Eidelheit 在波兰的 Lwow 工作，他于 1936 年在《Studia Mathematica》杂志上发表了多篇论文. 1939 年二战爆发，打断了他的工作(1940 年，他的论文发表在秘鲁的《Revista Ci.》和美国的《Annals of Mathematics》杂志上). 当 1948 年《Studia Mathematica》可以重新出版时，它收录了 Eidelheit 去世后发表的一篇论文，并说明他在 1943 年 3 月被杀. Dunford 和 Schwartz(1958，p. 460)介绍了分离定理的进一步发展. 在授予 Dunford 和 Schwartz 数学斯蒂尔奖时，Dunford(1981)说 Robert G. Bartle 写了大部分的历史注释和评论.

6.3 节 根据 Hardy、Littlewood 和 Polya(1934，1952)，"Jensen 建立了凸函数理论." Jensen (1906，p. 191)写到："Il me semble que la notion 'fonction convexe' est à peu près aussi fondamentale que celles-ci：'fonction positive'，'fonction croissante'. Si je ne me trompe pas en ceci la notion devra trouver sa place dans les expositions élémentaires de la théorie des fonctions réelles"有关近来更多的发展，例如，参见 Roberts 和 Varberg(1973)和 Rockafellar(1970).

6.4 节 L^p 的对偶是 L^q，其中 $1 \leqslant p < \infty$，$p^{-1}+q^{-1}=1$，(定理 6.4.1)当 μ 是$[0，1]$上的勒贝格测度时，这个结论首先是由 F. Riesz(1910)证明的.

6.5 节 接下来的注释是以 Dunford 和 Schwartz(1958，§ II. 5)为基础的. Hahn(1922)证明了巴拿赫空间上线性型序列的一致有界原理. 然后，Hildebrandt(1923)证明了更一般的情况. Banach 和 Steinhaus(1927)证明了只要 $T_\alpha x$ 在第二类范筹集中对 x 来说是有界的就足够了. 因此，这个定理被称为"Banach-Steinhaus"定理.

Schauder(1930)证明了巴拿赫空间中的开映射定理(6.5.2)，Banach(1932)证明了 F-空间的开映射定理. Banach(1931，1932)还证明了闭图形定理(6.5.3)并将这两个定理扩展到某个拓扑群中.

开映射定理和哈恩-巴拿赫定理在偏微分方程理论中是很有用的. (Hörmander，1964，p. 65，87，98，101).

6.6 节 Hadwiger 和 Ohmann 给出了 Brunn-Minkowski 不等式(6.6.1)的相关的初等证明，此外，还有一些其他的证明(Bonnesen 和 Fenchel，1934). 对 $k=2，3$，时，此不等式是由 Brunn(1887，Kap. III；1889，Kap. III)证明的. Minkowski(1910)证明了定理 6.6.1 中等式成立，如果 $V(A)>0$，$V(B)>0$，当且仅当 A 和 B 是位似的(对某个 $c \in \mathbb{R}$ 和 $v \in \mathbb{R}^k$，有 $B=cA+v$). 这些早期的参考文献

219 都是取自于 Bonnesen 和 Fenchel(1934，p. 90—91)的.

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到.

Banach, Stefan (1929). Sur les fonctionelles linéaires I, II. *Studia Math. 1*: 211–216, 223–239.

——— (1931). Über metrische Gruppen. *Studia Math. 3*: 101–113.

——— (1932). *Théorie des Opérations Linéaires*. Monografje Matematyczne I, Warsaw.

——— and Hugo Steinhaus (1927). Sur le principe de la condensation de singularités. *Fund. Math. 9*: 50–61.

Bonnesen, Tommy, and Werner Fenchel (1934). *Theorie der konvexen Körper*. Springer, Berlin. Repub. Chelsea, New York, 1948.

*Brunn, Hermann (1887). Über Ovale and Eiflächen. Inauguraldissertation, Univ. München.

*——— (1889). Über Kurven ohne Wendepunkte. Habilitationsschrift, Univ. München.

*——— (1910). Zur Theorie der Eigebiete. *Arch. Math. Phys.* (Ser. 3) *17*: 289–300.

Dieudonné, Jean (1941). Sur le théorème de Hahn-Banach. *Rev. Sci. 79*: 642–643.

——— (1943). Sur la séparation des ensembles convexes dans un espace de Banach. *Rev. Sci. 81*: 277–278.

Dunford, Nelson (1981). [Response]. *Amer. Math. Soc. Notices 28*: 509–510.

——— and Jacob T. Schwartz, with the assistance of William G. Badé and Robert G. Bartle (1958). *Linear Operators: Part I, General Theory.* Interscience, New York.

Eidelheit, Maks (1936). Zur Theorie der konvexen Mengen in linearen normierten Räumen. *Studia Math. 6*: 104–111.

Hadwiger, H., and D. Ohmann (1956). Brunn-Minkowskischer Satz und Isoperimetrie. *Math. Zeitschr. 66*: 1–8.

Hahn, Hans (1922). Über Folgen linearer Operationen. *Monatsh. für Math. und Physik 32*: 3–88.

——— (1927). Über lineare Gleichungssyteme in linearen Räumen. *J. für die reine und angew. Math. 157*: 214–229.

Hardy, Godfrey Harold, John Edensor Littlewood, and George Polya (1934). *Inequalities.* Cambridge University Press 2d ed., 1952, repr. 1967.

Hildebrandt, T. H. (1923). On uniform limitedness of sets of functional operations. *Bull. Amer. Math. Soc. 29*: 309–315.

Hörmander, Lars (1964). *Linear partial differential operators.* Springer, Berlin.

Jensen, Johan Ludvig William Valdemar (1906). Sur les fonctions convexes et les inégalites entre les valeurs moyennes. *Acta Math. 30*: 175–193.

Kelley, John L., Isaac Namioka, and eight co-authors (1963). *Linear Topological Spaces.* Van Nostrand, Princeton. Repr. Springer, New York (1976).

Kirszbraun, M. D. (1934). Über die zusammenziehenden und Lipschitzschen Transformationen. *Fund. Math. 22*: 77–108.

Lipschitz, R. O. S. (1864). Recherches sur le développement en séries trigonométriques des fonctions arbitraires et principalement de celles qui, dans un intervalle fini, admettent une infinité de maxima et de minima (in Latin). *J. reine angew. Math. 63*: 296–308; French transl. in *Acta Math. 36* (1913): 281–295.

McShane, Edward J. (1934). Extension of range of functions. *Bull. Amer. Math. Soc. 40*: 837–842.

Minkowski, Hermann (1897). Allgemeine Lehrsätze über die konvexen Polyeder. *Nachr. Ges. Wiss. Göttingen Math. Phys. Kl. 1897*: 198–219; *Gesammelte Abhandlungen, 2* (Leipzig and Berlin, 1911, repr. Chelsea, New York, 1967), pp. 103–121.

——— (1901). Sur les surfaces convexes fermées. *C. R. Acad. Sci. Paris 132*: 21–24; *Gesammelte Abhandlungen, 2*, pp. 128–130.

——— (1910). *Geometrie der Zahlen.* Teubner; Leipzig; repr. Chelsea, New York, 1953.

——— (1911, posth.). Theorie der konvexen Körper, insbesondere Begründung des Oberflächenbegriffs. *Gesammelte Abhandlungen, 2*, pp. 131–229.

Minty, George J. (1970). On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions. *Bull. Amer. Math. Soc. 76*: 334–339.

Riesz, F. (1910). Untersuchungen über Systeme integrierbarer Funktionen. *Math. Annalen 69*: 449–497.

Roberts, Arthur Wayne, and Dale E. Varberg (1973). *Convex Functions.* Academic Press, New York.

Rockafellar, R. Tyrrell (1970). *Convex Analysis.* Princeton University Press.

Schauder, Juliusz (1930). Über die Umkehrung linearer, stetiger Funktionaloperationen. *Studia Math. 2*: 1–6.

第7章 测度、拓扑与微分

在数学中用到的每一个测度几乎都是在空间上定义的，空间自身也是一种拓扑，其中测度的定义域或者是由拓扑生成的博雷尔σ-代数，关于测度的完备化或者是一个媒介的σ-代数．在一个测度空间上定义一个实值函数的积分并不涉及诸如定义域空间上的任何拓扑，尽管它是在值域空间$\mathbb{R}$上构造的（次序和拓扑是一样的）结构．7.1节将会探讨测度和拓扑之间的关系．

Radon-Nikodym定理(5.5节)定义了一个测度关于另一个测度的导数$\frac{d\gamma}{d\mu}=f$，这里假定γ关于μ是绝对连续的．一个很自然的问题就是当集合A缩小到$\{x\}$时，$\gamma(A)/\mu(A)$收敛到$f(x)$微分是否有意义．对此，显然，集合A包含x且它们的测度趋于0是不够的，因为大多数集合可能远离x．我们希望这些集合被包括在x的邻域中，构成收敛到x的一组滤子基．在$\mathbb{R}$中，对于普通微分，集合A是区间，通常在x处有一个端点．它对集合A收敛到x是不充分的．例如在$\mathbb{R}^2$中，如果集合A是包含x的矩形区域，矩形边长的比是无穷的，则仍然存在测度可微性的反例(关于勒贝格测度绝对连续)．7.2节将论述微分．

其他几节将会探讨测度和拓扑的更深层的关系．

7.1 贝尔σ-代数、博雷尔σ-代数和测度的正则性

对任意的拓扑空间$(X, \mathcal{T})$，博雷尔σ-代数$\beta(X, \mathcal{T})$定义为由$\mathcal{T}$生成的σ-代数．这个σ-代数中的所有集合称为博雷尔集．贝尔(Baire)σ-代数$Ba(X, \mathcal{T})$定义为最小的σ-代数，其中定义于此代数上的所有连续函数是可测的，并且值域空间为$\mathbb{R}$上的博雷尔σ-代数．

222

令$C_b(X, \mathcal{T})$是X上的所有有界连续实函数所形成的空间．对任意的实函数f，有界函数$\arctan f$是可测的，当且仅当f是可测的(对于f定义域上的任意σ-代数)，并且$\arctan f$是连续的，当且仅当f是连续的．因此对于$C_b(X, \mathcal{T})$中所有可测的f，$Ba(X, \mathcal{T})$也是最小的σ-代数．

显然，每一个贝尔集都是博雷尔集：$Ba(X, \mathcal{T}) \subset \mathcal{B}(X, \mathcal{T})$．下面将证明这两个$\sigma$-代数在度量空间中是等同的，但并不是在所有的空间中都是等同的．

7.1.1 定理 *在任意的度量空间(S, d)中，每一个博雷尔集都是贝尔集，所以对于度量拓扑$\mathcal{T}$，$Ba(S, \mathcal{T}) = \mathcal{B}(S, \mathcal{T})$．*

证明 设F为S上的任意闭集．令

$$f(x) := d(x,F) := \inf\{d(x,y): y \in F\}, \quad x \in S.$$

那么对S中任意的x和u，$|d(x, F) - d(u, F)| \leq d(x, u)$，如式(2.5.3)中所示．因此，$f$是连续的．因为$F$是闭的，$F = f^{-1}\{0\}$，所以$F$是一个贝尔集，从而它的补集也是贝尔集．结论得证．　□

例：在不可数多个紧的且每一个都多于一个点的豪斯多夫空间的笛卡儿积K中，一个单元素集是闭的，因此也是一个博雷尔集，但不是贝尔集，详述如下．由斯通-魏尔斯特拉斯定理(2.4.11)知，仅依赖有限多个坐标的连续实值函数所构成的集合$C_F(K)$在所有连续实函数所构成的集合$C(K)$中是稠密的．从而对$\forall f \in C(K)$，在$C_F(K)$中存在f_n，使得$\sup_{x \in K} |(f - f_n)(x)| < 1/n$．因此，$f$至多依赖于可数多个坐标，取与$f_n$有关的坐标集的并．这些贝尔集的集族仅依赖于可数多个坐标，

因此包括生成 σ-代数的集合，并且在取余和可数并下是闭的(对这些指标集重复取可数并)，这样就得到了整个贝尔 σ-代数，从而它不包含单元素集.

用一些比较容易处理的集合(如闭集或紧集)来逼近一般的可测集通常是有用的. 为此回顾闭集和紧集的关系是有帮助的：任意紧集的闭子集是紧的(定理2.2.2)；且在一个豪斯多夫空间中，任意紧集是闭的(命题2.2.9). 223

定义 令$(X, \mathcal{T})$是一个拓扑空间，μ 是某个 σ-代数$\mathcal{S}$上的测度. 那么$\mathcal{S}$中的集合 A 称为是**正则的**，如果

$$\mu(A) = \sup\{\mu(K): K\text{ 紧}, K \subset A, K \in \mathcal{S}\}.$$

如果 μ 是有限的，称 A 是**胎紧的**(tight)当且仅当 X 是正则的. 同样，A 称为是**闭正则的**当且仅当

$$\mu(A) = \sup\{\mu(F): F\text{ 闭}, F \subset A, F \in \mathcal{S}\}.$$

那么 μ 称为(闭)正则的当且仅当$\mathcal{S}$中的每个集合都是(闭)正则的.

例：$\mathbb{R}^k$ 的博雷尔 σ-代数上的任意有限测度是胎紧的，因为$\mathbb{R}^k$ 是紧集 $K_n := \{x: |x| \leqslant n\}$ 的可数并.

令 A 为无理数集$\mathbb{R} \setminus Q$，且对于任意的博雷尔集 $B \subset \mathbb{R}$，令 $\mu(B) := \lambda(B \cap [0, 1])$，其中 λ 是勒贝格测度. 为表明 A 是正则的，给定 $\varepsilon > 0$，令$\{q_n\}$为有理数的计数序列，对每个 n 都有相应的 q_n，取长度为 $\varepsilon/2^n$ 的开区间 U_n. 令 U 为所有 U_n 的并集，$K := [0, 1] \setminus U$. 那么 $K \subset A$ 且 $\mu(A \setminus K) < \varepsilon$. 令 $\varepsilon \downarrow 0$ 证明 A 是正则的.（这将会在证明定理7.1.3——$\mathbb{R}$ 中的所有博雷尔集都是正则的时得到证明.）

正则的概念将会经常用到，尽管在包括贝尔 σ-代数 $\mathcal{B}_a(X, \mathcal{T})$的 σ-代数$\mathcal{S}$中不总是会用到. 下面的事实将有助于证明某些 σ-代数中的所有集合是正则的.

7.1.2 命题 令$(X, \mathcal{T})$为豪斯多夫拓扑空间，$\mathcal{S}$是 X 的子集构成的 σ-代数，且 μ 是$\mathcal{S}$上的有限胎紧测度. 令

$$\mathcal{R} := \{A \in \mathcal{S}: \text{对于 }\mu, A\text{ 和 }X \setminus A\text{ 是正则的}\},$$

那么$\mathcal{R}$是一个 σ-代数. 如果把“正则”替换成“闭正则”，结论同样也成立.

证明 由定义，$A \in \mathcal{R}$当且仅当 $X \setminus A \in \mathcal{R}$. 设 A_1，A_2，… 为$\mathcal{R}$中的集合，$A := \bigcup_{n \geqslant 1} A_n$. 给定 $\varepsilon > 0$，考虑A_n 中的紧集 K_n 和 $X \setminus A_n$ 中的 L_n，其中对所有的 n，$\mu(A_n \setminus K_n) < \varepsilon/3^n$，$\mu((X \setminus A_n) \setminus L_n) < \varepsilon/2^n$. 于是对某个 $M < \infty$，$\mu\left(\bigcup_{1 \leqslant n \leqslant M} A_n\right) > \mu(A) - \varepsilon/2$. 令 $K := \bigcup_{1 \leqslant n < M} K_n$，则 K 是紧的，$K \subset A$，且 $\mu(K) \geqslant \mu(A) - \varepsilon$. 令 $L := \bigcap_{1 \leqslant n < \infty} L_n$. 那么 L 是紧的并且 $\mu((X \setminus A) \setminus L) \leqslant \sum_n \varepsilon/2^n = \varepsilon$. 这样 A 和 $X \setminus A$ 是正则的. 因为 μ 是胎紧的，且$\mathcal{R}$是 σ-代数，所以 $X \in \mathcal{R}$. 对于“闭正则”，因为闭集的有限并和可数交是闭的，所以结论同样成立.（因为 X 本身是闭的，因此，用“胎紧”并且“闭”来代替“紧”，也可以得到类似结论.） 224

□

贝尔 σ-代数上的测度称为贝尔测度. 博雷尔 σ-代数中的测度称为博雷尔测度. 在度量空间 S 中，如果 S 是正则的，则所有的博雷尔集也是正则的.

7.1.3 定理 在任意的度量空间(S, d)上，任意有限博雷尔测度 μ 是闭正则的. 如果 μ 是胎紧的，则它是正则的.

证明 令 U 是任意开集，F 是其补集. 令

$$F_n := \{x : d(x,F) \geqslant 1/n\}, \quad n = 1,2,\cdots$$

(与定理 7.1.1 和定理 4.2.2 证明中的类似)，所以 F_n 是闭的且其并是 U. 从而对于闭正则性，命题 7.1.2 中的 σ-代数$\mathcal{R}$包含所有的开集，因此包含所有的博雷尔集，所以 μ 是闭正则的.

如果 μ 是胎紧的，则给定 $\varepsilon>0$，考虑满足 $\mu(S\setminus K)<\varepsilon/2$ 的紧集 K. 对任意的博雷尔集 B，取闭集 $F\subset B$，$\mu(B\setminus F)<\varepsilon/2$. 令 $L:=K\cap F$. 于是 L 是紧的(定理 2.2.2)，$L\subset B$，且 $\mu(B\setminus L)<\varepsilon$，所以 B 是正则的，μ 是正则的.

如果度量空间是完备可分的，则 S 及所有的博雷尔集的正则性总成立.

7.1.4 Ulam 定理 *在任意完备可分度量空间(S, d)上，任意有限博雷尔测度是正则的.*

证明 为证明 μ 是胎紧的，令$\{x_n\}_{n\geqslant 1}$是 S 中的稠密序列. 对任意的 $\delta>0$ 和 $x\in S$，令 $\bar{B}(x, \delta):=\{y: d(x, y)\leqslant\delta\}$. 给定 $\varepsilon>0$，对每个 $m=1, 2, \cdots$，取 $n(m)<\infty$，使得

$$\mu\left(S\Big\backslash\bigcup_{n=1}^{n(m)}\bar{B}\left(x_n,\frac{1}{m}\right)\right)<\frac{\varepsilon}{2^m}.$$

令

$$K:=\bigcap_{m\geqslant 1}\bigcup_{n=1}^{n(m)}\bar{B}\left(x_n,\frac{1}{m}\right),$$

那么 K 在一个完备空间中是完全有界的且闭的，因此由命题 2.4.1 和定理 2.3.1 知，它是紧的. 而

$$\mu(S\setminus K)\leqslant\sum_{m=1}^{\infty}\varepsilon/2^m=\varepsilon.$$

225 所以 μ 是胎紧的，从定理 7.1.3 可推出这个定理. □

对于度量空间 S 上的所有有限博雷尔测度是正则的，S 并不一定是完备的：首先，回顾一些不完备的度量空间，例如，有通常度量的$(0, 1)$与完备的度量空间是同胚的(“拓扑完备的”，即定理 2.5.4). 另一方面，拓扑空间称为 σ-紧的，当且仅当它是可数多个紧集的并. 从定理 7.1.3 可以直接得到，σ-紧的度量空间上的一个有限的博雷尔测度总是正则的，尽管一些 σ-紧的空间(比如有理数空间$\mathbb{Q}$)甚至不是拓扑完备的(推论 2.5.6). Ulam 定理更关注于度量空间，如可分的、无穷维的、并非 σ-紧的巴拿赫空间，在一个度量空间上，$[0, 1]$上的非勒贝格可测子集不是足够“坏”的，例如，并非所有有限的博雷尔测度是正则的(见下面的习题 9).

在不可度量的空间中，(局部地)紧空间备受关注. 在给出下面这个事实之后，紧豪斯多夫空间中的正则性将会在 7.3 节进一步讨论.

7.1.5 定理 *对任意的紧豪斯多夫空间 K，K 上的任意有限贝尔测度 μ 是正则的.*

证明 令 f 为 K 上的任意连续实函数，F 为$\mathbb{R}$ 中的任意闭集，则 $f^{-1}(F)$是紧的贝尔集，因此也是正则的. 故$\mathbb{R}\setminus F$ 是闭集 F_n 的可数并，正如定理 7.1.3 证明中的那样. 因此

$$K\setminus f^{-1}(F)=\bigcup_n f^{-1}(F_n),$$

是紧集的可数并. 由于$f^{-1}(F)$这样的集合生成了贝尔 σ-代数，由命题 7.1.2 知，所有的贝尔集是正则的. 故 $K\setminus f^{-1}(F)$是正则的. □

习题

1. 对于有限测度空间$(X, \mathcal{S}, \mu)$，假设存在点 x_i 和数 $t_i>0$，使得对任意 $A\in\mathcal{S}$，$\mu(A)=\sum\{t_i: x_i\in A\}$. (于是 μ 是纯原子测度，如 3.5 节所定义的.) 对每个 $x\in X$，假设单元素集$\{x\}\in\mathcal{S}$. 证明：在$\mathcal{S}$中的每一个集合都

是正则的.

2. 令μ为可分度量空间S上的有限博雷尔测度. 假定对所有的$x\in S$, $\mu(\{x\})=0$. 证明：存在一个第一范畴集A(无处稠密集的可数并，见定理2.5.2)，使得$\mu(S\setminus A)=0$.

226

3. 若μ是拓扑空间S上的一个博雷尔测度，如果F是使得$\mu(S\setminus H)=0$的最小闭集H，则闭集F称为是μ的支集. 证明：如果S是可度量的且可分的，那么μ的支集总存在. [提示：利用命题2.1.4.]
4. 对可分度量空间中的任意闭集F，证明：存在具有支集F的有限博雷尔测度μ. [提示：应用2.1节习题5. 同习题1一样定义一个测度.]
5. 令(S, d)是完备可分的度量空间，其中S非空. 假设S没有孤立点，即对每个$x\in S$, x在$S\setminus\{x\}$的闭包中. 证明：在S上存在一个博雷尔测度μ，且$\mu(S)=1$，对所有的$x\in S$，有$\mu(\{x\})=0$. [提示：定义一个从$[0, 1]$映射到S的1－1的博雷尔可测函数f且令$\mu=\lambda\circ f^{-1}$.]
6. 证明：在两点集合X上存在一个(非豪斯多夫)拓扑，使得贝尔σ-代数和博雷尔σ-代数是不同的.
7. 集合X的子集族$\mathcal{L}$称为格点当且仅当$\varnothing\in\mathcal{L}$, $X\in\mathcal{L}$，且对$\mathcal{L}$中的任意A, B, $A\cup B\in\mathcal{L}$且$A\cap B\in\mathcal{L}$. 证明：对于$\mathcal{L}$中的A和B，所有集合$A\setminus B$组成的集族$\mathcal{D}$是一个半环(3.2节). 然后证明：对$\mathcal{L}$中的A_i和B_i，由格点集族$\mathcal{L}$生成的代数$\mathcal{A}$(包含最小的代数)是所有有限并$\bigcup_{1\leqslant i\leqslant n} A_i\setminus B_i$的集族，其中对不同的$i$，集$A_i\setminus B_i$可取作是不相交的. (在任意的拓扑空间中，所有开集的集族组成了一个格点，同所有闭集的集族类似.)
8. 集合的一个格点$\mathcal{L}$称为是一个σ-格点，当且仅当对任意的序列$\{A_n\}\subset\mathcal{L}$，有$\bigcup_{n\geqslant 1} A_n\in\mathcal{L}$和$\bigcap_{n\geqslant 1} A_n\in\mathcal{L}$. 对任意的豪斯多夫拓扑空间$(X, \mathcal{T})$, S的子集的σ-代数$\mathcal{S}$，和$\mathcal{S}$上的任意有限测度μ，证明：对于μ, $\mathcal{S}$中的所有正则集的集族是一个σ-格点，所有闭正则集的集族也是一个σ-格点. [提示：参见命题7.1.2的证明.]
9. 令A为一个对于$[0, 1]$上的勒贝格测度λ不可测的集合，由定理3.4.4知，外测度$\lambda^*(A)=1$. 定义集合B上的一个测度μ，通过$\mu(B)=\lambda^*(B)$, B是A的博雷尔子集，应用定理3.3.6证明：对于μ，正则可测集的集族不是一个代数. [提示：A是正则的吗?]
10. 令X是一个可数集. 证明：对于X上的某个拓扑$\mathcal{T}$, X的子集的每一个σ-代数$\mathcal{A}$都是博雷尔σ-代数. [提示：尝试$\mathcal{T}=\mathcal{A}$.]
11. 令$I:=[0, 1]$，证明：I上的所有连续实函数集$C[0, 1]$是$\mathbb{R}^I$上的博雷尔集，其中$\mathbb{R}^I$是I上的所有实函数的集合且具有积拓扑. [提示：证明$C[0, 1]$是闭集F_{mn}的可数并的可数交，其中$F_{mn}:=\{f:=|x-y|\leqslant 1/m$蕴涵了$|f(x)-f(y)|\leqslant 1/n\}$.]

227

12. 令μ是可分度量空间上的一个有限博雷尔测度. 证明：对μ的每一个原子(如3.5节所定义的)，有一个$x\in A$，使得$\mu(A)=\mu(\{x\})$. 进而证明：μ是纯原子当且仅当它有习题1中给出的形式，且μ是非原子测度当且仅当对所有的x, $\mu(\{x\})=0$.

*7.2 勒贝格微分定理

令λ表示勒贝格测度，$dx:=d\lambda(x)$. 第一个定理是关于勒贝格积分的基本定理形式. 在定理的经典形式中，被积函数g是一个连续函数，且不定积分的导数几乎处处等于g. 对于一个勒贝格可积函数，如无理数集的示性函数，这个等式不是处处都成立，而是几乎处处成立.

7.2.1 定理 如果一个函数g在区间$[a, b]$是勒贝格可积的，则对于λ-几乎所有的$x\in(a, b)$,

$$\frac{d}{dx}\int_a^x g(t)\,dt = g(x).$$

证明 将用到两个引理.

7.2.2 引理 令$\mathcal{U}$是$\mathbb{R}$中开区间的一个集族且其有界并为W. 那么对任意$t<\lambda(W)$，存在一有限不相交的子集族$\{V_1, \cdots, V_q\}\subset\mathcal{U}$，使得

$$\sum_{i=1}^{q}\lambda(V_i) > \frac{t}{3}.$$

证明 由正则性(定理7.1.4)，取一个紧集 $K\subset W$，满足 $\lambda(K)>t$. 然后取 U_1，…，$U_n\in\mathcal{U}$，使得 $K\subset U_1\cup\cdots\cup U_n$，次序为 $\lambda(U_1)\geqslant\lambda(U_2)\geqslant\cdots\geqslant\lambda(U_n)$. 令 $V_1:=U_1$. 对 $j=2$，3，…，从最小的 m 通过 $V_j:=U_m:=U_{m(j)}$ 递归地定义 V_j，使得对 $i<j$，U_m 与 V_i 不相交. 令 W_i 是一个与 V_i 有相同中心的区间，但长度是 V_i 的3倍. 那么对每个 $r\geqslant 2$，或者 U_r 是其中一个 V_j，或者对某个 $k<r$，$U_r\cap V_i=U_k$，所以 $\lambda(U_r)\leqslant\lambda(U_k)$，且 U_r 包含于 W_i(如图7-2所示).

228

因此，若 q 是使得 V_i 有定义的最大的 i，则

$$t<\lambda(K)<\lambda\left(\bigcup_{j=1}^{n}U_j\right)\leqslant\sum_{i=1}^{q}\lambda(W_i)=3\sum_{i=1}^{q}\lambda(V_i).\quad\square$$

W_i

$V_i = U_k$ U_r

图 7-2

7.2.3 引理 如果 μ 是区间 $[a,b]$ 上的一个有限博雷尔测度，A 是区间 $[a,b]$ 的博雷尔子集且 $\mu(A)=0$，则对 λ-几乎所有的 $x\in A$，$(\mathrm{d}/\mathrm{d}x)\mu([a,x])=0$.

证明 只需证明对 λ-几乎所有的 $x\in A$，$\lim\limits_{h\downarrow 0}\mu((x-h,x+h))/h=0$. 对 $j=1$，2，…，令

$$P_j:=\left\{x\in A:\limsup_{h\downarrow 0}\mu((x-h,x+h))/h>1/j\right\}.$$

这里 lim sup 能被限制为有理数 $h>0$，因为问题中 h 的函数关于 h 是左连续的. 还有函数 $1_{(x-h,x+h)}(y)$ 关于 x，h 和 y 是联合博雷尔可测的，因此由 Tonelli-Fubini 定理(及其证明)知，$\mu((x-h,x+h))$ 关于 x，h 是联合博雷尔可测的. 因此 lim sup 是可测的，且 P_j 是可测的.

给定 $\varepsilon>0$，取一开集 $V\supset A$ 且 $\mu(V)<\varepsilon$(由闭正则性知，这样的 V 是存在的). 对每个 $x\in P_j$，存在一个 $h>0$，使得 $(x-h,x+h)\subset V$ 且 $\mu((x-h,x+h))>h/j$. $(x-h,x+h)$ 这些区间覆盖了 P_j，所以由引理7.2.2知，对任意的 $t<\lambda(P_j)$，有有限的不交区间 J_1，…，$J_q\subset V$，使得

$$t\leqslant 3\sum_{i=1}^{q}\lambda(J_i)\leqslant 6j\sum_{i=1}^{q}\mu(J_i)\leqslant 6j\mu(V)\leqslant 6j\varepsilon.$$

因此 $\lambda(P_j)\leqslant 6j\varepsilon$. 令 $\varepsilon\downarrow 0$，则 $\lambda(P_j)=0$，$j=1$，2，…，且令 $j\to\infty$ 引理得证. $\square$

现在证明定理7.2.1，对每个有理数 r，令

229

$$g_r:=\max(g-r,0),\quad f_r(x):=\int_a^x g_r(t)\,\mathrm{d}t.$$

对 μ 应用引理7.2.3，这里对任意博雷尔集 E，$\mu:=\int_E g_r(t)\,\mathrm{d}\lambda(t)$，则对满足 $g(x)\leqslant r$ 的 λ-几乎所有的 x，有 $\mathrm{d}f_r(x)/\mathrm{d}x=0$. 令 B 为零测集中所有有理数 r 的并集，这里 $g(x)\leqslant r$，但 $\mathrm{d}f_r(x)/\mathrm{d}x\neq 0$. 于是 $\lambda(B)=0$. 当 $a\leqslant x<x+h\leqslant b$ 时，

$$\int_x^{x+h}g(t)\,\mathrm{d}t-rh=\int_x^{x+h}g(t)-r\mathrm{d}t\leqslant\int_x^{x+h}g_r(t)\,\mathrm{d}t,$$

所以

$$\frac{1}{h}\int_x^{x+h}g(t)\,\mathrm{d}t\leqslant r+\frac{1}{h}\int_x^{x+h}g_r(t)\,\mathrm{d}t.$$

同样，如果 $a\leqslant x-h<x\leqslant b$，

$$\frac{1}{h}\int_{x-h}^{x}g(t)\,\mathrm{d}t\leqslant r+\frac{1}{h}\int_{x-h}^{x}g_r(t)\,\mathrm{d}t.$$

对于从包含一个点 x 的开区间映射到$\mathbb{R}$ 的函数 f，上、下、左、右的导数分别定义如下：

$$UR(f,x) := \lim_{h\downarrow 0}\sup(f(x+h)-f(x))/h \geqslant$$
$$LR(f,x) := \lim_{h\downarrow 0}\inf(f(x+h)-f(x))/h,$$
$$UL(f,x) := \lim_{h\downarrow 0}\sup(f(x)-f(x-h))/h \geqslant$$
$$LL(f,x) := \lim_{h\downarrow 0}\inf(f(x)-f(x-h))/h.$$

这些量总是可定义的，但可能为 $+\infty$ 或 $-\infty$. 现在 f 在 x 处有(有限)导数，当且仅当四个导数相等(且有限).

如果 x 不在 B 中且 $r>g(x)$，则有 $\mathrm{d}f_r(x)/\mathrm{d}x=0$. 接下来，对 $f(x):=\int_a^x g(t)\,\mathrm{d}t$，令 $h\downarrow 0$，我们有 $UR(f,\ x)\leqslant r$ 且 $UL(f,\ x)\leqslant r$. 令 $r\downarrow g(x)$，有 $UR(f,\ x)\leqslant g(x)$ 且 $UL(f,\ x)\leqslant g(x)$. 考虑 $-g$，同样有 $UR(-f,\ x)\leqslant -g(x)$ 和 $UL(-f,\ x)\leqslant -g(x)$，所以 $LL(f,\ x)\geqslant g(x)$ 且 $LR(f,\ x)\geqslant g(x)$. 对于这样的 x，这四个导数都等于 $g(x)$，所以有

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x g(t)\,\mathrm{d}t = g(x),$$

这就证明了定理 7.2.1. □

定理 7.2.7 将会证明另一个几乎处处可导的函数(尽管它们不可能是它们的不定积分)是如下定义的有界变差函数. 对任意的 $J\subset\mathbb{R}$ 上的实值函数 f，它在 J 上的全变差定义为：

230

$$\mathrm{var}_f J := \sup\left\{\sum_{i=1}^{n}|f(x_i)-f(x_{i-1})| : x_0 < x_1 < \cdots < x_n, x_i \in J\right\},$$

其中对 $i=0,\ 1,\ \cdots,\ n$，$x_i\in J$. 如果 $\mathrm{var}_f J<+\infty$，则 f 称为 J 上的有界变差函数. J 通常为闭区间 $[a,\ b]$.

有界变差函数被描述为不减函数的差分.

7.2.4 定理 $[a,\ b]$上的实函数是有界变差函数，当且仅当 $f=F-G$ 其中 F 和 G 是$[a,\ b]$上的不减实函数.

证明 如果 F 是不减的(记为“$F\uparrow$”)，即对 $a\leqslant x\leqslant y\leqslant b$，$F(x)\leqslant F(y)$，则显然 $\mathrm{var}_F[a,\ b]=F(b)-F(a)<+\infty$. 如果 $G\uparrow$，$\mathrm{var}_{F-G}[a,\ b]\leqslant\mathrm{var}_F[a,\ b]+\mathrm{var}_G[a,\ b]<+\infty$.

反之，如果 f 是$[a,\ b]$上的任意有界变差实函数，令 $F(x):=\mathrm{var}_f[a,\ x]$. 那么显然 $F\uparrow$. 令 $G:=F-f$，则对 $a\leqslant x\leqslant y\leqslant b$，

$$G(y)-G(x) = F(y)-F(x)-(f(y)-f(x)) \geqslant 0$$

因为 $\mathrm{var}_f[a,\ x]+f(y)-f(x)\leqslant\mathrm{var}_f[a,\ y]$. 所以 $G\uparrow$ 且 $f=F-G$. □

对任意可数集 $\{q_n\}\subset\mathbb{R}$，可以由 $F:=\sum\{2^{-n}: q_n\leqslant t\}$ 定义不减函数 F，且 $0\leqslant F\leqslant 1$. 于是 F 是不连续的且在每个 q_n 处跳跃的高度为 2^{-n}，这里 $\{q_n\}$ 可以是有理数的稠密集. 这表明任意可数集可以是不连续集.

7.2.5 定理 如果 f 是$[a,\ b]$上的有界变差实函数，则至多除了一个可数集外，f 是连续的.

证明 由定理 7.2.4，我们可以假设 $f\uparrow$. 那么对 $a<x\leqslant b$，有 $\lim f(x^-):=\lim_{u\uparrow x}f(u)$ 且对 $a\leqslant x<b$，有 $f(x^+):=\lim_{v\downarrow x}f(v)$. 如果 $a\leqslant u<x<v\leqslant b$，那么有

$$\mathrm{var}_f[a,b] \geqslant |f(x)-f(u)|+|f(v)-f(x)|.$$

令 $u\uparrow x$ 且 $v\downarrow x$，有

$$\mathrm{var}_f[a,b] \geqslant |f(x)-f(x^-)|+|f(x^+)-f(x)|.$$

类似地，对任意有限集 $E\subset(a, b)$，有

$$\mathrm{var}_f[a,b] \geqslant \sum_{x\in E} |f(x^+)-f(x)|+|f(x)-f(x^-)|.$$

从而对 $n=1, 2, \cdots$，至多有有限个 $x\in[a, b]$，满足

$$|f(x^+)-f(x)|>1/n \text{ 或 } |f(x)-f(x^-)|>1/n.$$

231 因此，至多除去一个可数集，我们有 $f(x^+)=f(x)=f(x^-)$，所以 f 在 x 处连续. □

回顾 f 在 x 是右连续的当且仅当 $f(x^+)=f(x)$（如在 3.1 节和 3.2 节中），然而 $1_{[0,1)}$ 是右连续的，但上述情况对 $1_{[0,1]}$ 不成立. 对任意有限非负测度 μ，$a\in\mathbb{R}$，$f(x):=\mu((a, x])$ 给出了一个不减右连续函数 f. 相反地，给定这样一个 f，由定理 3.2.6 知，有这样一个 μ. 更确切地说，测度 μ 和函数 f 之间的关系能被推广到符号测度和有界变差函数.

7.2.6 定理 *给定 $[a, b]$ 上的一个实函数 f，下面两个条件等价：*

(i) 在博雷尔子集 $[a, b]$ 上存在一个可数可加有限的符号测度 μ，对 $a\leqslant x\leqslant b$，有 $\mu((a, x])=f(x)-f(a)$.

(ii) f 是 $[a, b)$ 上的有界变差右连续函数.

证明 如果 (i) 成立，由可数可加性，f 在 $[a, b)$ 上是右连续的. 由哈恩–若尔当分解（定理 5.6.1），可以假设 μ 是一个非负有限测度. 但是 f 是不减的，所以它的全变差是 $f(b)-f(a)<+\infty$，正如所要求的.

如果 (ii) 成立，则由定理 7.2.4，对某个不减的 g 和 h，有 $f=g-h$，且 g，h 的这个结论蕴涵着 f 是不减的，所以可以假设 f 是不减的，则由定理 3.2.6 的结论成立. □

在讲连续性之前，先来回顾一个例子，在命题 4.2.1 的证明中定义的康托尔函数 g，它是一个从 $[0, 1]$ 映上到自身的不减连续函数，在每个"中间第三个"区间上为一个常数：在 $[1/3, 2/3]$ 上，$g=1/2$，在 $[1/9, 2/9]$，$g=1/4$，等等. 那么对 $1/3<x<2/3$，$1/9<x<2/9$，$7/9<x<8/9$ 等，有 $g'(x)=0$. 因此，在 $[0, 1]$ 上对于勒贝格测度几乎处处有 $g'(x)=0$，但是 g 不是定理 7.2.1 中自身的导数的不定积分，因为尽管 g 是连续的，但它不是一个常数.（物理学家大都忽略这样的函数和测度，直到在其上出现"怪引子"他们才定义，参见 Grebogi et al.，1987.）

下面是勒贝格理论中的另一个主要结论. "几乎处处"（"a.e."）意味着除了一个勒贝格零测集外都成立.

7.2.7 定理 *对区间 (a, b) 上的任意有界变差实函数 f，导数 $f'(x)$ 几乎处处存在，且在 (a, b) 上的*

232 *勒贝格测度集 $\mathcal{L}^1$ 中.*

证明 对 $a<x<b$，令 $g(x):=f(x^+)$，$h(x):=g(x)-f(x)$，则由定理 7.2.5 知，除了可数集 C 中的 x 外，有 $h(x)=0$. 令 $V:=\mathrm{var}f(a, b)$. 对任意有限的 $F\subset C$，$x\in F$，取很接近 x 的 $y_x>x$，使得区间 $[x, y_x]$ 是不相交的，故 $\sum_{x\in F}|f(y_x)-f(x)|\leqslant V$. 对每个 $x\in F$ 令 $y_x\downarrow x$，有 $\sum_{x\in F}|h(x)|\leqslant V$. 然后令 $F\uparrow C$，有 $\sum_{x\in C}|h(x)|\leqslant V$，所以 h 和 $g\equiv f+h$ 是有界变差函数，只需证明关于 g 和 h 的定理.

首先，对于 h，令 $v(A):=\sum_{x\in A\cap C}|h(x)|$ 为定义在 (a, b) 的所有博雷尔子集 A 上的有限测度. 因为 v 集中在可数集 C 上，所以对任意的 $s\in(a, b)$，对 $x\in B$，有 $(\mathrm{d}/\mathrm{d}x)v([s, x])=0$，这里 $\lambda((s, b)\setminus B)=0$. 如果在 (a, b) 中 $s<t<u<v$，那么 $|h(v)-h(u)|/(v-u)\leqslant v((u, v])/(v-u)$，

如果 $h(u)=0$，或者 $h(u)\leqslant v((t,v])/(v-u)$，如果 $h(v)=0$. 若 $v\in B$，令 $t\uparrow v$，或者若 $u\in B$，令 $v\downarrow u$，在 B 上得到 $h'=0$. 令 $s\downarrow a$，则关于 λ-几乎处处于 (a,b) 上有 $h'=0$.

所以只需证明定理对 g 成立，换句话说，当 f 是右连续的. 仍由定理 7.2.4，可以假设 $f\uparrow$. 我们可以通过设置 $f(a):=f(a^+)$，$f(b):=f(b^-)$ 把 f 扩张到 $[a,b]$ 而不改变其全变差. 令 μ 为 $[a,b]$ 的博雷尔集上的测度，满足 $\mu((a,x])=f(x)-f(a)$，$a\leqslant x\leqslant b$（且 $\mu(\{a\})=0$），由定理 7.2.6 给出. 由勒贝格分解（定理 5.3.3），我们有 $\mu=\mu_{ac}+\mu_{sing}$，其中 μ_{ac} 是绝对连续的，μ_{sing} 关于勒贝格测度是奇异的. 对 $a<x\leqslant b$，令 $g(x):=\mu_{ac}((a,x])$，$h(x):=\mu_{sing}((a,x])$. 于是由引理 7.2.3 得，$h'(x)=0$ a. e.

由 Radon-Nikodym 定理（5.5.4）知，存在一个函数 $j\in\mathcal{L}^1([a,b],\lambda)$，有

$$g(x)=\int_a^x j(t)\,\mathrm{d}t,\quad a<x\leqslant b.$$

因此，由定理 7.2.1 知，$g'(x)$ 几乎处处存在且等于 $j(x)$，且 $f'(x)=j(x)$ a. e. □

区间 $[a,b]$ 上的函数 f 称为是绝对连续的，当且仅当对每个 $\varepsilon>0$，存在 $\delta>0$，使得对 $[a,b]$ 的任意不交子区间 $[a_i,b_i)$ $(i=1,2,\cdots)$，$a_i<b_i$ 且 $\sum_i(b_i-a_i)<\delta$，有 $\sum_i|f(b_i)-f(a_i)|<\varepsilon$.

前四个问题都与函数的绝对连续性有关. 结合定理 7.2.1 与定理 7.2.7 的证明，表明 $[a,b]$ 上的函数 f 是绝对连续的，当且仅当对于某个关于勒贝格测度绝对连续的符号测度 μ，$f(x)-f(a)\equiv\mu((a,x])$.

233

习题

1. 证明：对任意绝对连续函数是有界变差函数. ［提示：如果 f 有无界变差，它就不会在任意的小区间上绝对连续.］
2. 假设习题 1 成立. 令 f 为有界变差函数，在 $[a,b]$ 上右连续. 证明：f 是绝对连续的，当且仅当定理 7.2.6 定义的符号测度 μ 关于勒贝格测度是绝对连续的. ［提示：与 5.5 节习题 4 比较.］
3. 假设习题 1 和习题 2 成立. 证明：f 在 $[a,b]$ 上是绝对连续的，当且仅当对某 $j\in\mathcal{L}^1([a,b],\lambda)$，$f(x)=f(a)+\int_a^x j(t)\,\mathrm{d}t$，$x\in[a,b]$. ［提示：这是 Radon-Nikodym 定理（5.5.4）的前身.］
4. 证明：如果 f 和 g 在 $[a,b]$ 上是绝对连续的，则它们的乘积 fg 也是绝对连续的.
5. 设 $\{J_i\}_{1\leqslant i\leqslant n}$ 是 $\mathbb{R}$ 上有界开区间构成的一个有限集族. 证明：存在一个由不交区间组成的子集族，其总长度至少为 $\lambda\left(\bigcup_i J_i\right)/2$. 利用此结论来给出引理 7.2.2 的另一种证明. ［提示：我们可以假设其中一个 J_i 是空的. 令每个非空的 $J_i=(a_i,b_i)$. 选取 V_1 为由 a_i 的最小值及 b_i 的最大值构成的区间. 令 $V_j=(c_j,d_j)$，从 $j=1$ 开始，用递归的方法给定 V_j，选取 V_{j+1} 作为一个 j_i，如果存在一个 i，使得 $d_i\in J_i$，且在 j_i 中，有最大的值 b_i. 如果这样的 i 不存在，令 $V_{j+1}:=\varnothing$，且令 V_{j+2} 为一个 J_i，如果存在一个这样的 J_i，有 $a_i\geqslant d_j$，且在这个 J_i 中，有最小的 a_i 的那个也有最大的 b_i. 如果这样的 J_i 不存在，构造结束. 选取不交区间族或者是 V_1，V_3，…或者是 V_2，V_4，…这里可以去掉空的区间.］
6. 设 $\mathcal{S}$ 为 $\mathbb{R}$ 的开子区间族，且其并集为 U. 假设对每个 $\varepsilon>0$，U 中的每一个点都包含在 $\mathcal{S}$ 的一个长度小于 ε 的开区间中. 证明：存在一个由不交区间构成的序列 $\{V_n\}\subset\mathcal{S}$，使得并集 $V:=\bigcup_n V_n$ 几乎是所有的 U，换句话说，$\lambda(U\setminus V)=0$. ［提示：利用习题 5 及迭代.］
7. 设 $\{D_i\}_{1\leqslant i\leqslant n}$ 为 $\mathbb{R}^2$ 中开圆盘的任意有限族，$D_i:=B(p_i,r_i):=\{z\in\mathbb{R}^2:|z-p_i|<r_i\}$ 其中 $r_i>0$，$p_i\in\mathbb{R}^2$，且 $|\cdot|$ 是 $\mathbb{R}^2$ 中一个向量的一般欧几里得长度. 证明：存在互不相交 D_i 的集族 D，其并集 V 有面积

$\lambda^2(V) \geqslant \lambda^2\left(\bigcup_i D_i\right)/4$. [提示：用归纳法. 取一个有最大半径的 D_i，把它置于 D 中，去掉所有与它相交的 D_j，且把这个归纳假设应用到其他集合中.]

234

8. 设$\mathcal{E}$为$\mathbb{R}^2$ 中的开圆盘族，其并集为 U，有对每个 $x\in U$，$\mathcal{E}>0$，有一个半径小于$\mathcal{E}$包含 x 的圆盘 $D\in\mathcal{E}$. 证明：存在一个$\mathcal{E}$的不交子集族$\mathcal{D}$，其并集为 V，使得 $\lambda^2(U\setminus V)=0$. [提示：参见习题 6 和习题 7.]
9. 把习题 5 ~ 8 扩展到$\mathbb{R}^d$，其中 $d\geqslant 3$，有常量 $1/2^d$.
10. 设 μ 为$\mathbb{R}^k$ 上的测度，关于勒贝格测度 λ^k 绝对连续，Radon-Nikodym 导数为 $f=\mathrm{d}\mu/\mathrm{d}\lambda^k$. 证明：对 λ^k-几乎所有的 x，任意 $\varepsilon>0$，存在一个 $\delta>0$，使得对于任意包含 $x(|x-p|<r)(r<\delta)$ 的球 $B(p,r)$，有 $|f(x)-\mu(B(p,r))/\lambda^k(B(p,r))|<\varepsilon$.

*7.3 正则性扩张

回顾拓扑空间$(S,\mathcal{T})$称为局部紧的，当且仅当 S 中的每个点 x 有一个紧的邻域(即对某个开 V 和紧 L，$x\in V\subset L$). 一些紧的豪斯多夫不可度量空间的主要例子是不可数多个紧的豪斯多夫空间的乘积(吉洪诺夫定理 2.2.8). 对于关于局部紧但不可度量空间的测度理论，下面的定理是基本的. 本节的后续部分会对它进行证明. 这个定理不适用于度量空间，其中所有的博雷尔集是贝尔集(定理 7.1.1).

7.3.1 定理 *令 K 为一紧的豪斯多夫空间且 μ 为 K 上的任意有限的贝尔测度. 那么 μ 能扩张为 K 上的博雷尔测度，且有一个唯一正则的博雷尔扩张.*

证明 对任意紧集 L，令 $\mu^*(L):=\inf\{\mu(A):A\supset L\}$，其中 A 取遍整个贝尔集.

7.3.2 引理 *对任意不相交的紧集 L 和 M，*

$$\mu^*(L\cup M)=\mu^*(L)+\mu^*(M).$$

证明 我们知道一个紧豪斯多夫空间是正规的(定理 2.6.2)，且任意两个不相交的紧集可由一连续实函数分离(定理 2.6.3). 从而存在不相交的贝尔集 A 和 B，有 $A\supset L$ 和 $B\supset M$. 那么对任意的贝尔集 $C\supset L\cup M$，

$$\mu(C)\geqslant\mu(C\cap A)+\mu(C\cap B)\geqslant\mu^*(L)+\mu^*(M),$$

235

所以 $\mu^*(L\cup M)\geqslant\mu^*(L)+\mu^*(M)$. 相反的不等式也成立(见引理 3.1.5)，故等式成立. □

现对任一博雷尔集 B，令

$$\nu(B):=\sup\{\mu^*(L):L\text{ 紧},L\subset B\}.$$

如果 B 是紧的，显然有 $\nu(B)=\mu^*(B)$. 如果 $B\subset C$，那么 $\nu(B)\leqslant\nu(C)$. 令$\mathcal{M}(\nu)$为所有博雷尔集 B 构成的集族，使得对所有的博雷尔集 A 有 $\nu(A)=\nu(A\cap B)+\nu(A\setminus B)$. 由引理 7.3.2，我们总有 $\nu(A)\geqslant\nu(A\cap B)+\nu(A\setminus B)$，所以对所有的博雷尔集 A，

7.3.3 $$B\in\mathcal{M}(\nu)\text{ 当且仅当 }\nu(A)\leqslant\nu(A\cap B)+\nu(A\setminus B).$$

类似地在(7.3.3)中，我们能用“紧集”代替“博雷尔集”，因为一方面，所有的紧集是博雷尔集；另一方面，对所有的紧集 $L\subset A$，如果 $\nu(L)\leqslant\nu(L\cap B)+\nu(L\setminus B)$，则因为 $\mu^*(L)=\nu(L)$，由 ν 的定义给出了(7.3.3)中的不等式. 现在由引理 3.1.8 的证明得$\mathcal{M}(\nu)$是一个代数且 ν 在其上是有限可加的.

7.3.4 引理 *对任意闭集 $L\subset K$，有 $L\in\mathcal{M}(\nu)$.*

证明 我们需要证明对任意紧集 $M\subset K$，

$$\mu^*(M)\leqslant\mu^*(M\cap L)+\sup\{\mu^*(N):N\subset M\setminus L,N\text{ 紧}\}.$$

给定 $\varepsilon>0$，取一满足 $\mu(V)<\mu^*(M\cap L)+\varepsilon$ 的贝尔集 $V\supset M\cap L$. 由定理 7.1.5，我们可以取 V 为开集. 那么 $M\setminus V$ 是紧的，且包含在 $M\setminus L$ 中，所以

$$\mu^*(M\cap L)+\nu(M\setminus L)\geqslant\mu(V)-\varepsilon+\mu^*(M\setminus V).$$

对任意贝尔集 $W\supset M\setminus V$，$M\subset V\cup W$，因此 $\mu^*(M)\leqslant\mu(V)+\mu(W)$. 令 $\mu(W)\downarrow\mu^*(M\setminus V)$ 且 $\varepsilon\downarrow 0$ 就得到所证结论. □

7.3.5 引理 *$\mathcal{M}(\nu)$ 是所有博雷尔集的 σ-代数且 ν 是其上的测度.*

证明 我们知道 $\mathcal{M}(\nu)$ 是一个代数，又由引理 7.3.4 知，它包含所有的开集、闭集和紧集. 剩下来只需要检验单调收敛性.

首先，令 B_n 为博雷尔集，$B_n\downarrow\varnothing$. 假设 $\nu(B_n)\geqslant\varepsilon>0$. 取紧集 $K_n\subset B_n$ 且 $\mu^*(K_n)\geqslant\nu(B_n)-\varepsilon/3^n$. 令 $L_n:=\bigcup_{1\leqslant j\leqslant n}K_j$. 因为 L_n 和 K_j 在 $\mathcal{M}(\nu)$ 中是紧的，对所有的 n，我们有 $\nu(B_n\setminus K_n)\leqslant\varepsilon/3^n$. 又因为 [236]

$$B_n\setminus L_n\subset\bigcup_{i\leqslant j\leqslant n}B_j\setminus K_j,$$

对所有的 n，有 $\nu(B_n\setminus L_n)\leqslant\varepsilon/2$，且 $\nu(L_n)\geqslant\varepsilon/2>0$. 所以 $L_n\neq\varnothing$. 但是 $L_n\subset B_n\downarrow\varnothing$ 蕴涵了 $L_n\downarrow\varnothing$，与紧性矛盾：$K\subset\bigcup_n K\setminus L_n$ 没有有限子覆盖. 因此 $\nu(B_n)\downarrow 0$.

现在令 $E_n\in\mathcal{M}(\nu)$ 且 $E_n\uparrow E$. 我们想要证明 $E\in\mathcal{M}(\nu)$ 且 $\nu(E_n)\uparrow\nu(E)$. 因为 E 是博雷尔集，对每个 n，$\nu(E)=\nu(E_n)+\nu(E\setminus E_n)$，且 $E\setminus E_n\downarrow\varnothing$，因此 $\nu(E\setminus E_n)\downarrow 0$，正如刚刚所证明的. 因此 $\nu(E_n)\uparrow\nu(E)$.

对任意一个博雷尔集 A 和所有的 n，

$$\nu(A)=\nu(A\cap E_n)+\nu(A\setminus E_n)\leqslant\nu(A\cap E)+\nu(A\setminus E_n).$$

令 $D_n:=A\setminus E_n\downarrow D:=A\setminus E$. 我们想要证明 $\nu(D_n)\downarrow\nu(D)$. 显然，对某个 $\alpha\geqslant\nu(D)$，有 $\nu(D_n)\downarrow\alpha$. 如果 $\alpha>\nu(D)$，取 $0<\varepsilon<\alpha-\nu(D)$. 对每个 $n=1,2,\cdots$ 取一个紧集 $C_n\subset D_n$ 满足 $\mu^*(C_n)>\nu(D_n)-\varepsilon/3^n$. 令 $F_n:=\bigcap_{1\leqslant j\leqslant n}C_j$. 则对每个 n，

$$C_n\subset F_n\cup\bigcup_{1\leqslant i\leqslant n}C_n\setminus C_i.$$

因为所有的紧集在代数 $\mathcal{M}(\nu)$ 中，

$$\nu(C_n)\leqslant\nu(F_n)+\sum_{1\leqslant i\leqslant n}\nu(C_n\setminus C_i).$$

因为对 $i\leqslant n$，$C_n\subset D_n\subset D_i$，

$$\nu(C_n)\leqslant\nu(F_n)+\sum_{1\leqslant i\leqslant n}\varepsilon/3^i.$$

所以 $\nu(F_n)\geqslant\nu(C_n)-\varepsilon/2\geqslant\alpha-\varepsilon$. 令 $F:=\bigcap_{1\leqslant n<\infty}F_n=\bigcap_{1\leqslant n<\infty}C_n$，则 $F\subset D$，且 F 是紧集，所以 $\nu(D)\geqslant\mu^*(F)$. 因为 F_n 和 F 在 $\mathcal{M}(\nu)$ 中，且 $F_n\downarrow F$，由证明的第一部分，我们有 $F_n\setminus F\downarrow\varnothing$ 且 $\nu(F_n)\downarrow\nu(F)$. 但是 $\mu^*(F)=\nu(F)\geqslant\alpha-\varepsilon>\nu(D)$，矛盾. 所以 $\nu(D_n)\downarrow\nu(D)$，且 $\nu(A)\leqslant\nu(A\cap E)+\nu(D)$，故 $E\in\mathcal{M}(\nu)$. 一个博雷尔集 $E\in\mathcal{M}(\nu)$ 当且仅当 $K\setminus E\in\mathcal{M}(\nu)$. 因此，如果 $E_n\in\mathcal{M}(\nu)$ 且 $E_n\downarrow E$，则 $E\in\mathcal{M}(\nu)$. 所以 $\mathcal{M}(\nu)$ 是一个单调类，由定理 4.4.2 知，$\mathcal{M}(\nu)$ 是一个 σ-代数且 ν 是其上的一个测度，这就证明了引理 7.3.5. □

现在回到定理 7.3.1 的证明，对任意一个贝尔集 B 和紧集 $L\subset B$，有 $\mu(B)\geqslant\mu^*(L)$，所以 $\mu(B)\geqslant\nu(B)$. 同样，$\mu(K\setminus B)\geqslant\nu(K\setminus B)$，所以 $\mu(B)\leqslant\nu(B)$，则 $\mu(B)=\nu(B)$. 所以 ν 扩张为 μ. [237]

因为在紧集上 $\nu=\mu^*$，由定义知，ν 是正则的.

现在证明唯一性，设 ρ 为对于博雷尔集 μ 的另一个正则扩张. 那么对于任意紧集 L，$\rho(L)\leqslant\mu^*(L)$，因此对任意的博雷尔集 B，由正则性得，$\rho(B)\leqslant\nu(B)$. 同样 $\rho(K\setminus B)\leqslant\nu(K\setminus B)$，且 $\rho(K)=\nu(K)=\mu(K)$，故 $\rho(B)=\nu(B)$，这就证明了定理 7.3.1. □

习题

1. 假设 μ 是 X 的子集构成的代数 $\mathcal{A}$ 上的有限可加的非负测度，满足 $\mu(X)<\infty$. 令 $(X,\mathcal{F})$ 是豪斯多夫拓扑空间. 假定 μ 对于每个 $A\in\mathcal{A}$ 是正则的，$\mu(A)=\sup\{\mu(K):K\subset A, K$ 是紧集$\}$. 证明：μ 在 $\mathcal{A}$ 上是可数可加的. [提示：利用定理 3.1.1.]
2. 设 $(K,\leqslant)$ 是一个不可数良序集，并且有最大元素 Ω，使得对任意 $x<\Omega$，$\{y:y\leqslant x\}$ 是可数的. 在 K 上取区间拓扑，其子基由 $\{\{x:x<\beta\}:\beta\in K\}\cup\{\{x:x>\alpha\}:\alpha\in K\}$ 给出. 那么 K 就是一个紧的豪斯多夫空间. 对于任意的博雷尔集 A，如果 A 包含一个 $\{x:x<\Omega\}$ 中相对闭的不可数集，则令 $\mu(A)=1$，否则，令 $\mu(A)=0$. 证明：μ 是一个非正则测度且没有支集.（称 F 是 μ 的一个支集，当且仅当 F 是 μ 的最小闭子集且其补集的测度为 0.）[提示：如果 C 和 D 是不可数闭集，证明 $C\cap D$ 是不可数的，即对于任意的 $a\in K$，取 $a<c_1<d_1<c_2<d_2<\cdots$，$c_i\in C$，$d_i\in D$，则 $\sup c_i=\sup d_i\in C\cap D$. 令 $\mathcal{C}:=\{A:A$ 或 $K\setminus A$ 包含一个不可数闭集$\}$. 利用单调类(定理 4.4.2)证明 $\mathcal{C}$ 包含所有博雷尔集. 利用定理 3.1.1 证明 μ 是可数可加的.]
3. 证明：存在一个紧的豪斯多夫空间 X，且它不是任何有限正则的博雷尔测度的支集. [提示：对于除去 x_0 外的所有 x，令 X 是具有 $\{x\}$ 的不可数开集. 设 x_0 的邻域是包含 x_0 和几乎所有的其他有限点的集合. 那么 X 是紧的. 证明 X 的子集 A 是 X 上的一个有限正则博雷尔测度的支集，当且仅当 A 是可数的，若无限，则包含 x_0.]
4. 令 $I:=[0,1]$，$S:=2^I$ 有积拓扑. 证明：S 中的贝尔 σ-代数 $\mathcal{A}$ 不是任何拓扑 $\mathcal{F}$ 的 σ-代数. [提示：假定结论不成立. 回顾定理 7.1.1 证明后面的例子，下面是证明的梗概：
   (a) 对于每个 $x\in S$，令 D_x 是 $\{x\}$ 对于 $\mathcal{T}$ 的闭包. 那么对于某个可数集 $C(x)$，如果对于所有的 $t\in C(x)$，有 $y(t)=x(t)$，则 $y\in D_x$.

   238

   (b) 设 $x\leqslant y$ 当且仅当 $y\in D_x$，证明这是一个偏序，且 $x\leqslant y$ 当且仅当 $D_y\subset D_x$.
   (c) 通过递归，对于 $\leqslant$，定义集合 $W\subset S$ 是一个良序集，这里对于任意的 $w\in W$，$\{v\in W:v\leqslant w\}$ 是可数的. 证明 W 可取为不可数的，且对于 $x\in W$，D_x 的交集对于 $\mathcal{T}$ 是闭集，但不在 $\mathcal{A}$ 中，矛盾.]
5. 在引理 7.3.5 的证明过程中，可以发现对于任意博雷尔集列 $B_n\downarrow\varnothing$，我们有 $v(B_n)\downarrow 0$. 在那个点处应用定理 3.3.1，就可以直接证明出结论吗？请解释.

*7.4 $C(K)$ 的对偶和傅里叶级数

对于任意紧的豪斯多夫空间 K，令 $C(K)$ 表示 K 上的所有连续实函数组成的线性空间，且上确界范数为 $\|f\|_\infty:=\sup\{|f(x)|:x\in K\}$. 因为所有的连续函数在 K 上均有界，所以由定理 2.4.9 知，$(C(K),\|\cdot\|_\infty)$ 是巴拿赫空间.

令 $(X,\mathcal{S})$ 是一个可测空间，f 是 X 上的一个实可测函数，μ 是 $\mathcal{S}$ 上一个符号测度. 我们取若尔当分解 $\mu=\mu^+-\mu^-$（定理 5.6.1）. 对于任意实可测函数 f，定义 $\int f\mathrm{d}\mu:=\int f\mathrm{d}\mu^+-\int f\mathrm{d}\mu^-$，无论何时，这都是有定义的，使得如果 $\int f^+\mathrm{d}\mu^+=+\infty$ 或 $\int f^-\mathrm{d}\mu^-=+\infty$，那么 $\int f^-\mathrm{d}\mu^+$ 和 $\int f^+\mathrm{d}\mu^-$ 是有限的.

特别地，若 $(X,\mathcal{T})$ 是一个拓扑空间，并且 $C_b(X,\mathcal{T})$ 是 X 上的有界连续实函数组成的空间，$f\in C_b(X,\mathcal{T})$，μ 是 X 上的有限有符号的贝尔测度，则 $\int f\mathrm{d}\mu$ 总是有定义的并且有限的. 令 $I_\mu(f):=\int f\mathrm{d}\mu$，

则 I_μ 属于对偶巴拿赫空间 $C_b(X, \mathcal{T})'$. 例如，如果 μ 是一个点式群体 δ_x，且对于任意(贝尔)集 A，$\delta_x(A) := 1_A(x)$，则对于任意函数 f，有 $I_\mu(f) = f(x)$.

L^p 空间的对偶空间的元素表示(定理 6.4.1)通常称为“Riesz 表示”定理，于是有如下定理.

7.4.1 定理 对于任意紧的豪斯多夫空间 X，$\mu \mapsto I_\mu$ 是 X 上所有有限有符号的贝尔测度空间 $M(X, \mathcal{T})$ 的线性等距同构，范数为 $\|\mu\| := |\mu|(X)$，满射于 $C(X)'$上，且具有对偶范数 $\|\cdot\|'_\infty$，其中 $\|T\|'_\infty := \sup\{|T(f)| : \|f\|_\infty \leq 1\}$.

证明 显然 $\mu \mapsto I_\mu$ 是一个从 $M(X, \mathcal{T})$ 映射到 $C(X)'$的线性映射. 对于给定的 μ，根据定理 5.6.1，取哈恩分解 $X = A \cup (X \setminus A)$，其中 $\mu^+(X \setminus A) = \mu^-(A) = 0$，这里 A 是一个贝尔集. 根据贝尔测度的正则性(定理 7.1.5)，对于给定的 $\varepsilon > 0$，取紧集 $K \subset A$ 和 $L \subset X \setminus A$，其中 $\mu^+(A \setminus K) < \varepsilon/4$，$\mu^-((X \setminus A) \setminus L) < \varepsilon/4$. 根据 Urysohn 引理(2.6.3)，取 $f \in C(X)$ 且对于任意的 x，$-1 \leq f(x) \leq 1$，在 K 上 $f = 1$，在 L 上 $f = -1$. 故 $\|f\|_\infty \leq 1$ 且 $\left|\int f \mathrm{d}\mu\right| \geq |\mu|(X) - \varepsilon$. 令 $\varepsilon \downarrow 0$，得出 $\|I_\mu\|' = |\mu|(X)$，因此 $\mu \mapsto I_\mu$ 是一个等距同构.

[239]

为证明它是映上的，令 $L \in C(X)'$. 与证明定理 6.4.1 相同，根据引理 6.4.2，我们有 $L = L^+ - L^-$，这里 L^+ 和 L^- 都属于 $C(X)'$，并且对于 $C(X)$ 中所有的 $f \geq 0$，有 $L^+(f) \geq 0$ 和 $L^-(f) \geq 0$. 由 4.5 节的定义，显然 $C(X)$ 是一个斯通向量格. 则根据迪尼定理(2.4.10)，L^+ 和 L^- 是准整数. 因而根据 Stone-Daniell 定理(4.5.2)，对任意的 $f \in C(X)$，有测度 ρ 和 σ，使得 $L^+(f) = \int f \mathrm{d}\rho, L^-(f) = \int f \mathrm{d}\sigma$. 因为常量 $1 \in C(K)$，所以这里的 ρ 和 σ 是有限测度. 那么对所有的 $f \in C(X)$，有 $\mu := \rho - \sigma \in M(X, \mathcal{T})$ 和 $L(f) = \int f \mathrm{d}\mu$，这就证明了定理 7.4.1. □

根据正则性扩张(定理 7.3.1)，可以取 μ 为一个正则的博雷尔测度. 上述定理也经常被表述为另一种形式.

本节其余部分我们将讨论傅里叶级数与一致有界性原理之间的关系. 令 S^1 是一个单位圆 $\{z : |z| = 1\} \subset \mathbb{C}$，这里复平面$\mathbb{C}$是在附录 B 中有定义. 令 $C(S^1)$ 是 S^1 上的所有连续复值函数组成的巴拿赫空间，其中上确界范数为 $\|\cdot\|_\infty$. 根据复数域上的斯通-魏尔斯特拉斯定理(推论 2.4.13)，幂 $z \mapsto z^m (m \in \mathbb{Z})$ 的有限线性组合在 $C(S^1)$ 中对于 $\|\cdot\|_\infty$(一致收敛)是稠密的. 设 μ 是 S^1 上的自然旋转不变量概率测度，也就是，$\mu = \lambda \circ e^{-1}$，这里 $e(x) := \mathrm{e}^{2\pi i x}$，$\lambda$ 是$[0, 1]$上的勒贝格测度. 我们也可以记作 $\mathrm{d}\mu := \mathrm{d}\theta/(2\pi)$，这里 θ 是单位圆中通常的角坐标，$0 \leq \theta \leq 2\pi (\theta = 2\pi x)$. 对于任意测度空间$(X, \mu)$和 $1 \leq p < \infty$，回顾定义为 $L^p(X, \mu, K)$ 的 L^p 空间，这里数域$K = \mathbb{R}$ 或$\mathbb{C}$.

7.4.2 命题 $\{z^m\}_{m \in \mathbb{Z}}$构成了 $L^2(S^1, \mu, \mathbb{C})$空间的一组规范正交基.

证明 根据定理 5.2.1，L^2 空间是完备的(希尔伯特空间). 容易验证对于 $m \in \mathbb{Z}$，z^m 是规范正交的. 如果它们不是一组基，则根据定理 5.3.8 和定理 5.4.9，存在$f \in \mathcal{L}^2(S^1, \mu)$，且对于所有的 $m \in \mathbb{Z}$ 和 $\|f\|_2 > 0$，有$f \perp z^m$. 由定理 5.5.2，有 $L^2 \subset L^1$. 令$\mathcal{A}$为幂 z^m 的所有有限线性组合的集合，则对任意的 $g \in C(S^1)$，存在 $P_n \in A$ 且 $\|P_n - g\|_\infty \to 0$. 从而 $\int f \cdot (P_n - g) \mathrm{d}\mu \to 0$. 因为对所有的 n，$\int f P_n \mathrm{d}\mu = 0$(注意到共轭复数$\overline{P_n} \in \mathcal{A}$)，故 $\int f g \mathrm{d}\mu = 0$. 对于任意的博雷尔集 $A \subset S^1$，令 $\nu(A) := \int_A f \mathrm{d}\mu$.

[240]

那么由定理7.4.1推得，$\| I_\nu \|'$是ν的全变差，显然有$\int |f|\,\mathrm{d}\mu > 0$，但是$I_\nu = 0$，产生矛盾，命题得证. □

函数$f\in\mathcal{L}^1(S^1,\mu)$的傅里叶级数定义为级数$\sum_{n\in\mathbb{Z}} a_n z^n$，其中$a_n := a_n(f) := \int f(z) z^{-n}\mathrm{d}\mu(z)$. 这个级数在不同的条件下将以不同的形式收敛. 例如，如果$f\in\mathcal{L}^2$，根据命题7.4.2和规范正交基(定理5.4.7上面)的定义，它的傅里叶级数在$\mathcal{L}^2$中以范数收敛到f. 然而，对于一致收敛性，情况要复杂得多.

7.4.3 命题 *存在$f\in C(S^1)$，使得f的傅里叶级数在$z=1$点不收敛于f，并且不一致收敛于f.*

证明 设$S_n(f)$表示f在1处的傅里叶级数的前n项和，即

$$S_n(f) := \sum_{m=-n}^{n} \int f(z) z^{-m}\mathrm{d}\mu(z) = \int f(z) D_n(z)\mathrm{d}\mu(z),$$

其中$D_n(z) := \sum_{m=-n}^{n} z^m = (z^{n+1} - z^{-n})/(z-1), z\neq 1$，因为对每个$m$，$z^m + z^{-m}$是实值的，所以此式是实值的. 于是$S_n$是$C(S^1)$上的连续线性型，根据定理7.4.1，$\| S_n \|'$等于由$\nu(A) = \int_A D_n(z)\mathrm{d}\mu(z)$所定义的符号测度$\nu$的全变差. 这个全变差的值是$\int |D_n|\,\mathrm{d}\mu(z)$. 然后，根据一致有界性原理(6.5.1)，只需证明当$n\to\infty$时，$\| D_n \|_1 \to \infty$. 由变量变换$z = \mathrm{e}^{\mathrm{i}x}$，我们要证

$$\int_{-\pi}^{\pi} |(\mathrm{e}^{\mathrm{i}(n+1)x} - \mathrm{e}^{-\mathrm{i}nx})/(\mathrm{e}^{\mathrm{i}x} - 1)|\,\mathrm{d}x \to \infty$$

或等价地，对分子分母同除以$\mathrm{e}^{\mathrm{i}x/2}$，

$$\int_0^{\pi} \left|\sin\left(\left(n+\frac{1}{2}\right)x\right)\right| \Big/ \sin\left(\frac{x}{2}\right)\mathrm{d}x \to \infty.$$

易知$\sin(nx + x/2) = \sin(nx)\cos(x/2) + \cos(nx)\sin(x/2)$. 对于$\theta := x/2$，我们有$0\leqslant\theta\leqslant\pi/2$，所以$\cos^2\theta\leqslant\cos\theta$，$1 - 2\cos\theta + \cos^2\theta \leqslant \sin^2\theta$，且$|1-\cos\theta| \leqslant \sin\theta$. 所以只需证明

$$\int_0^{\pi} |\sin(nx)| \Big/ \sin\left(\frac{x}{2}\right)\mathrm{d}x \to \infty.$$

241 为此，我们可以利用如下引理.

7.4.4 引理 *对于区间$[a,b]$上的任意连续实函数g，有*

$$\lim_{n\to\infty}\int_a^b |\sin(nx)|\, g(x)\,\mathrm{d}x = \frac{2}{\pi}\int_a^b g(x)\,\mathrm{d}x,$$

证明 设对于$a\leqslant x\leqslant b$，$|g(x)|\leqslant 1$. 给定$\varepsilon>0$，取足够小的$\delta>0$，当$|x-y|<\delta$时，有$|g(x)-g(y)|<\varepsilon$. 然后取足够大的n，使得$2\pi/n<\delta$. 对于这样的n，我们可以把区间$[a,b]$分解为互不相交且每段长度均为$2\pi/n$的区间$I(j) := [a_j, b_j)$，其余区间I_0的长度为$\lambda(I_0) < 2\pi/n$. 当$n\to\infty$时，I_0上的积分趋近于0. 对每个j，

$$\left|\int_{I(j)} |\sin(nx)|\,(g(x) - g(a_j))\,\mathrm{d}x\right| \leqslant 2\pi\varepsilon/n.$$

上式对所有的j求和不超过$(b-a)\varepsilon$. 对于每个j，$|\sin(nx)|$在$I(j)$上的平均值等于$|\sin t|$在长度为2π(一个完全周期)的区间上的平均值，这个平均值是$2/\pi$. 引理得证. □

以下，继续证明命题7.4.3，我们可得对任意$c>0$，当$n\to\infty$时，

$$\int_{c}^{\pi}\frac{|\sin(nx)|}{\sin(x/2)}\mathrm{d}x\to\frac{2}{\pi}\int_{c}^{\pi}\frac{1}{\sin(x/2)}\mathrm{d}x.$$

因为当 $x\to 0$ 时，$(\sin x)/x\to 1$，$\int_0^1 1/x\mathrm{d}x=+\infty$，所以可令 $c\downarrow 0$ 从而得到 $\|D_n\|_1\to\infty$，命题得证. □

习题

1. 设在 S^1 上 $\mu=\delta_1-\delta_i+\mathrm{i}\delta_{-1}$. 求在什么条件下满足 $\|f\|_\infty\leqslant 1$ 的一个复值函数 $f\in C_b(S^1)$ 可以使等式 $\int f\mathrm{d}\mu=\|I_\mu\|'$ 成立？
2. 设 K 是任意一个紧的豪斯多夫空间. 证明：K 上的连续函数构成的空间 $C(K)$ 在空间 $L^p(K,\mu)$ 中是稠密的，其中的 p 满足 $1\leqslant p\leqslant\infty$，$\mu$ 是 K 上的任意有限贝尔测度. [提示：如果结论不成立，利用 6.1 节的习题 4、定理 6.4.1 和定理 7.4.1.]
3. 证明：存在由函数 $x\mapsto a_n\sin(nx)(n\geqslant 1)$ 和 $x\mapsto b_n\cos(nx)(n\geqslant 0)$ 构成的空间 $L^2([0,2\pi),\lambda,\mathbb{R})$ 的一组规范正交基，并且对于所有的 n，求 a_n，b_n.
4. 设 S 是一个局部紧的豪斯多夫空间. 设 $C_0(S)$ 是 S 上所有连续实函数构成的且有紧支撑的空间. 设 L 是 $C_0(S)$ 上的线性函数，满足当 $f\geqslant 0$ 时，$L(f)\geqslant 0$. 证明：存在一个如 5.6 节习题 4～5 定义的拉东测度 μ，对于所有的 $f\in C_0(S)$，有 $L(f)=\int f\mathrm{d}\mu$.

242

5. 设 $\alpha=\nu+\mathrm{i}\rho$ 是 S^1 上的一个有限复值测度，其中 ν 和 ρ 是有限的实值符号测度. 定义 α 的傅里叶级数为(形式)级数 $\sum_{m\in\mathbb{Z}}a_m z^m$，其中 $a_m:=\int z^{-m}\mathrm{d}\alpha(z)$. 称一个有限复测度序列 α_n 弱收敛于测度 α，如果对于所有的 $f\in C(S^1)$，当 $n\to\infty$ 时，$\int f\mathrm{d}\alpha_n\to\int f\mathrm{d}\alpha$. 回顾 $\mathrm{d}\mu(\theta)=\mathrm{d}\theta=2\pi$，对每个博雷尔集 $A\subset S^1$，任意函数 $g\in L^1(S^1,\mu)$ 定义一个复测度 $[g]$ 为 $[g](A):=\int_A g\mathrm{d}\mu$，对于有限复测度 α，令 $\alpha_n:=[\sum_{|m|\leqslant n}a_m a^m]$，证明：如果对某个 $h\in\mathcal{L}^2(S^1,\mu)$，$\alpha=[h]$，那么 α_n 弱收敛到 α，但若对任意 $w\in S^1$ 有 $\alpha=\delta_w$，则结论不成立. [提示：参考命题 7.4.3 的证明.]
6. 证明：当 $n\to\infty$ 时，复测度 $[z^n]$ 弱收敛(弱收敛的定义见习题 5)，并且求出它的极限. [提示：参考引理 7.4.4 的证明.]
7. 令 $(X,\mathcal{T})$ 是一个拓扑空间，μ 和 ν 是关于 $\mathcal{T}$ 的贝尔 σ-代数 $\mathcal{S}$ 上的两个有限测度. 设 $1\leqslant p\leqslant\infty$，在什么条件下 $C_b(X,\mathcal{T})$ 上的线性型 $I_\nu:f\mapsto\int f\mathrm{d}\nu$ 对于半范为 $\left(\int|f|^p\mathrm{d}\mu\right)^{1/p}$ 的 $\mathcal{L}^p(\mu)$ 的拓扑是连续的. (根据勒贝格分解和 Radon-Nikodym 定理给出 μ 和 ν 应满足的条件.)

*7.5 几乎一致收敛和 Lusin 定理

可测函数不一定处处连续，例如，有理数集合的示性函数. 这个函数在测度为 0 的点集的补集上是连续的. 又如，给定 $\varepsilon>0$，在第 n 个有理数邻域内取长为 $\varepsilon/2^n$ 的开区间. 这些区间的并是一个测度小于 ε 的稠密开集，它的示性函数不是连续的，甚至除去任何测度为 0 的集合，它也不是连续的. 但是，除去一个测度足够小的集合，这个函数却是连续的. 对于更一般的可测函数，此结论将由(定理 7.5.2)给出. 首先，证明除去足够小的集合，逐点收敛是一致收敛的.

7.5.1 定理(Egoroff 定理) *设 $(X,\mathcal{S},\mu)$ 是有限测度空间. 设 f_n 和 f 是从 X 映射到以 d 为度量的可分度量空间 S 的可测函数. 假定对 μ-几乎所有的 x，有 $f_n(x)\to f(x)$，则对任意 $\varepsilon>0$，存在满足 $\mu(X\setminus A)<\varepsilon$ 的集合 A，使得在 A 上 f_n 一致收敛于 f，即*

243

$$\lim_{n\to\infty}\sup\{d(f_n(x),f(x)):x\in A\}=0.$$

证明 对于 m, $n=1, 2, \cdots$令

$$A_{mn}:=\{x:\text{对所有的 } k\geqslant n, d(f_k(x),f(x))\leqslant 1/m\}.$$

由命题4.1.7知，每一个 A_{mn} 都是可测的，则对于每一个 m，当 $n\to\infty$ 时，$\mu(X\setminus A_{mn})\downarrow 0$. 选取 $n(m)$，使得 $\mu(X\setminus A_{mn(m)})<\varepsilon/2^m$. 令 $A:=\bigcap_{m\geqslant 1}A_{mn(m)}$，则在 A 上 $\mu(X\setminus A)<\varepsilon$ 且 f_n 一致收敛于 f. □

例如，函数 x^n 在$[0, 1)$上几乎处处收敛于0，并且在区间$[0, 1-\varepsilon)(\varepsilon>0)$上一致收敛. 回忆拓扑空间的博雷尔 σ-代数上的一个有限测度 μ 称为是闭正则的，如果对于每一个博雷尔集合 B，$\mu(B)=\sup\{\mu(F): F\text{ 闭}, F\subset B\}$. 由定理7.1.3可知，一个度量空间上的任意有限博雷尔测度都是闭正则的.

下面的定理是 Lusin 将关于实变量实函数的定理扩张到更一般的定义域和值域空间.

7.5.2 定理(Lusin 定理) 设$(X, \mathcal{T})$是任意拓扑空间，μ 是 X 上有限的闭正则博雷尔测度. 设(S, d)为一个可分度量空间，f 是从 S 映射到 X 的博雷尔可测函数，则对任意 $\varepsilon>0$，存在一个闭集 $F\subset X$，使得 $\mu(X\setminus F)<\varepsilon$ 且 f 对于 F 的限制是连续的.

证明 设$\{s_n\}_{n\geqslant 1}$是 S 的可数稠密子集. 对于 $m=1, 2, \cdots$和任意 $x\in X$，设对使得 $d(f(x), s_n)<1/m$ 成立的最小的 n，有 $f_m(x)=s_n$，则 f_m 是定义在 X 上的可测函数. 对于每一个 m，设 $n(m)$足够大，使得

$$\mu\{x:\text{对所有的 } n\leqslant n(m), d(f(x),s_n)\geqslant 1/m\}\leqslant 1/2^m.$$

根据闭正则性，对 $n=1, \cdots, n(m)$，取一个闭集 $F_{mn}\subset f_m^{-1}\{s_n\}$，使得

$$\mu(f_m^{-1}\{s_n\}\setminus F_{mn})<\frac{1}{2^m n(m)}.$$

对于每一个固定的 m，集合 F_{mn}对不同的 n 值是互不相交的. 令 $F_m:=\bigcup_{n=1}^{n(m)}F_{mn}$，则 f_m 在 F_m 上是连续的. 通过选择 $n(m)$和 F_{mn}可得，$\mu(X\setminus F_m)<2/2^m$.

因为 $d(f_m, f)<1/m$ 处处成立，显然 f_m 一致收敛于 f(因此不需要 Egoroff 定理7.5.1). 对 $r=1, 2, \cdots$，令 $H_r:=\bigcap_{m=r}^{\infty}F_m$，则 H_r 是一个闭集且 $\mu(X\setminus H_r)\leqslant 4/2^r$. 取足够大的 r，使得 $4/2^r<\varepsilon$，则对 $m\geqslant r$，$H_r\subset F_m$ 上的连续函数 f_m 的一致极限 f 对于 H_r 是连续的. 所以可以取 $F=H_r$，定理得证. □

244

习题

1. 设$(X, \mathcal{T})$是一个规范拓扑空间，μ 是 X 的贝尔 σ-代数上的有限闭正则测度. 设 $C_b(X, \mathcal{T})$是 X 关于$\mathcal{T}$的有界连续复函数空间，对每一个 $g\in L^1(X, \mu)$和 $f\in C_b(X, \mathcal{T})$，令 $T_g(f):=\int gf\mathrm{d}\mu$. 证明：$T_g\in C_b(X, \mathcal{T})'$和 $\|T_g\|'=\int|g|\mathrm{d}\mu$.
2. 如果存在一个开集 U 和属于第一范畴(无处稠密集的可数并)的集合 B 和 C 满足 $A=(U\setminus B)\cup C$，则称 A 在拓扑空间上具有贝尔性质. 证明：所有具有贝尔性质的集合形成的集类是一个 σ-代数，且包含博雷尔 σ-代数. [提示：不能直接作为可数交集处理.]
3. 证明：对任意从可分度量空间 S 映射到$\mathbb{R}$ 的博雷尔可测函数 f，存在一个属于第一范畴的集合 D，使得 f 对于 $S\setminus D$ 是连续的. [提示：利用习题2的结论.](这是 Lusin 定理的一个"相似范畴"，一定意义上讲，在范畴情况下比在测度情况下更强.)

4. 设$(X, \mathcal{S}, \mu)$是一个有限测度空间. 假设对某个$M<\infty$，对所有n和x，$\{f_n\}$是满足$|f_n(x)|\leqslant M$的可测函数序列. 利用 Egoroff 定理证明：$\int f_n \mathrm{d}\mu \to \int f \mathrm{d}\mu$（有界收敛定理——勒贝格控制收敛定理的特殊情况）.
5. 设f是康托尔集的一个示性函数. 定义一个满足$\lambda(F)>7/8$的特殊闭集$F\subset[0, 1]$，使得f对于F的限制是连续的.
6. 当f是无理数集合的示性函数时，证明同样的结论.

注释

7.1 节 博雷尔集现在一般定义为由拓扑生成的σ-代数上的一个集合，而以前则定义为一个局部紧空间中，由紧集生成的σ-环上的一个集合(Halmos，1950，p. 219)，相应地也定义了贝尔集.

Ulam 定理是由 Oxtoby 和 Ulam(1939)提出的，人们将它主要归功于 Ulam，原因是他计划独立发表一篇论文(准备发表在《Comptes Rendus Acad. Sci. Paris》上)，但实际上 Ulam 并没有完成这篇论文，他只是写了一个自传(1976).

245

J. von Neumann(1932)证明了一个比较弱的结论，即在定理 7.1.4 的证明中，不仅选取从 1 到$n(m)$的并，同时也选取了所有m的交集，所以K是闭的，且是直径小于等于$2/m$的集合的有限并(集合A的直径定义为$\sup\limits_{x,y\in A} d(x, y)$).

在没有出版的演讲稿(我从 Institute for Advanced Study 那里获得了一份复印件)中，von Neumann(1940—1941，p. 85—91)提出了正则测度的概念. Halmos(1950，p. 292，295)引用了 1940—1941 的注解作为正则测度的原始来源.

7.2 节 Lebesgue(1940，p. 120—125)证明了定理 7.2.1，关于勒贝格积分的微积分基本定理. 证明是在 Rudin(1974，p. 162—165)的基础上给出的，包括引理 7.2.2 和 7.2.3. Vitali(1904—1905)证明了一个函数是绝对连续的当且仅当它是$\mathcal{L}^1$函数的不定积分. 把有界变差函数分解成为两个非递减函数的差(定理 7.2.4)归功于 Camille Jordan(参见 5.6 节的注释). Lebesgue(1904，p. 128)证明了有界变差函数的几乎处处可微性(定理 7.2.7). 在证明的过程中，没有利用测度理论(除了 0 测度集的概念，它是陈述的要素)，参考 Riesz(1930—1932)与 Riesz and Nagy(1953，p. 3—10).

一些问题中的关于$\mathbb{R}^k$上的绝对连续测度的可微性是由 Vitali(1908)和 Lebesgue(1910)解决的.

7.3 节 Halmos(1950，§§ 51—54)论述了正则性扩张，参考文献中的 Ambrose(1946)、Kakutani and Kodaira(1944)、Kodaira(1941)和 von Neumann(1940—1941)从不同的方面论述了这一理论的发展. 感谢 Dorothy Maharam Stone 提供给我一些关于这个问题的实例.

7.4 节 根据 Riesz and Nagy(1953，p. 110)，定理 7.4.1 关于$C[0, 1]$这个实例的论述是由 F. Riesz(1909，1914)解决的. 定理新近的一些证明可以参考 Garling(1986).

S^1上的$\mathcal{L}^2$函数的傅里叶级数几乎处处收敛(对于μ)可以通过 L. Carleson(1966)的一个比较难的定理解决. R. A. Hunt(1968)证明了如果对于某个$p>1$有$f\in\mathcal{L}^p(S^1, \mu)$，或更一般地，如果

$$\int |f(z)|\,(\max(0,\log|f(z)|))^2 \mathrm{d}\mu(z) < \infty,$$

则傅里叶级数几乎处处收敛. C. Fefferman(1971)把 Carleson 的结果扩张到具有二元函数上(在$S^1\times S^1$上，即在环面T^2上).

Kolmogorov(1923，1926)发现$\mathcal{L}^1$中的函数的傅里叶级数若几乎处处(μ)发散，则函数几乎处处发散. Zygmund(1959，p. 306—310)给出了一个说明.

命题 7.4.3 首先是由 du Bois-Reymond(1876)证明的，当然是用一种不同的方法.

傅里叶级数最早是由傅里叶于 1807 年在一部关于热力学的论文集中提出的，这篇文章最终递交给法兰西学院并出版，有注释的版本在 Grattan-Guinness(1972)出版. 傅里叶称“la résolution d'une fonction arbitraire en sinus ou en cosinus d'arcs multiples”(Grattan-Guinness, 1972, p. 193). 这个级数

246

可以表示不连续函数；傅里叶提出：

$$\cos(u) - \frac{1}{3}\cos(3u) + \frac{1}{5}\cos(5u) - \cdots = \begin{cases} \pi/4, & \text{对 } |u| < \pi/2 \\ 0, & \text{对 } u = \pm\pi/2 \\ -\pi/4, & \text{对 } \pi/2 < |u| < 3\pi/2, \text{等等.} \end{cases}$$

另一方面，傅里叶的计算对于某些自变量的范围使用了泰勒级数. 傅里叶的论文集由 Laplace、Lagrange、Monge 和 S. F. Lacroix 审查，但他的观点并没有被接受. 显然，Lagrange 特别反对在此论述中用三角级数表示任意函数的不精确做法(Grattan-Guinness, 1972, p. 24). 这个学院于 1811 年提出把热在固体中的传播作为数学研究的重要课题，并加以评奖. 评定委员会由 Lagrange、Laplace、Legendre、R. J. Haüy 和 E. Malus(Herivel, 1975, p. 156)组成. 傅里叶获得了这个奖，获奖文章和他进一步的工作(Fourier, 1822)最终都被出版(Fourier, 1824, 1826). 关于傅里叶的介绍也可参考 Ravetz and Grattan-Guinness(1972).

7.5 节 定理 7.5.1 和定理 7.5.2 对实变量实函数的结论分别已由 Egoroff(1911)和 Lusin(1912)出版. 这些参考文献来源于 Saks(1937). 关于 Egoroff，请参考 Paplauskas(1971)，关于 Lusin，请参考 13.2 节的注释.

据我所知，对可测函数取值于任意可分度量空间和闭的有限正则测度空间的 Lusin 定理，首先是由 Schaerf(1947)给出的. Schaerf 证明了 f 取值于次可数拓扑空间(即一个关于空间的拓扑有一组可数基的空间)和更一般的定义域空间(“邻域空间”)的结论. 对于结论的扩张，请参考 Schaerf

247

(1948)和 Zakon(1965).

参 考 文 献

带星号的资料在次级来源中被讨论过，但在原著中并没见到.

*Ambrose, Warren (1946). *Lectures on topological groups* (unpublished). Ann Arbor.

*du Bois-Reymond, P. (1876). Untersuchungen über die Convergenz und Divergenz der Fourierschen Darstellungsformeln. *Abh. Akad. München 12*: 1–103.

Carleson, Lennart (1966). On convergence and growth of partial sums of Fourier series. *Acta Math. 116*: 135–157.

Egoroff, Dmitri (1911). Sur les suites de fonctions mesurables. *C. R. Acad. Sci. Paris 152*: 244–246.

Fefferman, Charles (1971). On the convergence of multiple Fourier series. *Bull. Amer. Math. Soc. 77*: 744–745.

Fourier, Jean Baptiste Joseph (1822). *Théorie analytique de la chaleur*. F. Didot, Paris.

*——— (1824). Théorie du mouvement de la chaleur dans les corps solides. *Mémoires de l'Acad. Royale des Sciences 4* (1819–1820; publ. 1824): 185–555.

*——— (1826). Suite du mémoire intitulé: “Théorie du mouvement de la chaleur dans les corps solides.” *Mémoires de l'Acad. Royale des Sciences 5* (1821–1822; publ. 1826): 153–246; *Oeuvres de Fourier, 2*, pp. 1–94.

——— (1888–1890, posth.) *Oeuvres*. Ed. G. Darboux. Gauthier-Villars, Paris.

Garling, David J. H. (1986). Another ‘short’ proof of the Riesz representation theorem. *Math. Proc. Camb. Phil. Soc. 99*: 261–262.

Grattan-Guinness, Ivor (1972). *Joseph Fourier, 1768–1830*. MIT Press, Cambridge, Mass.

Grebogi, Celso, Edward Ott, and James A. Yorke (1987). Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. *Science 238*: 632–638.

Halmos, Paul (1950). *Measure Theory*. Van Nostrand, Princeton.

Herivel, John (1975). *Joseph Fourier: The Man and the Physicist*. Clarendon Press, Oxford.

Hunt, R. A. (1968). On the convergence of Fourier series. In *Orthogonal Expansions and Their Continuous Analogues*, pp. 235–255. Southern Illinois Univ. Press.

*Kakutani, Shizuo, and Kunihiko Kodaira (1944). Über das Haarsche Mass in der lokal bikompakten Gruppe. *Proc. Imp. Acad. Tokyo 20*: 444–450.

*Kodaira, Kunihiko (1941). Über die Beziehung zwischen den Massen und Topologien in einer Gruppe. *Proc. Math.-Phys. Soc. Japan 23*: 67–119.

Kolmogoroff, Andrei N. [Kolmogorov, A. N.] (1923). Une série de Fourier-Lebesgue divergente presque partout. *Fund. Math. 4*: 324–328.

——— (1926). Une série de Fourier-Lebesgue divergente partout. *C. R. Acad. Sci. Paris 183*: 1327–1328.

Lebesgue, Henri Léon (1904). Leçons sur l'intégration et la recherche des fonctions primitives, Paris. 2d ed., 1928. Repr. in *Oeuvres Scientifiques 2*, pp. 111–154.

——— (1910). Sur l'intégration des fonctions discontinues. *Ann. Ecole Norm. Sup.* (Ser. 3) *27*: 361–450. Repr. in *Oeuvres Scientifiques 2*, pp. 185–274.

Lusin, Nikolai (1912). Sur les propriétés des fonctions mesurables. *C. R. Acad. Sci. Paris 154*: 1688–1690.

von Neumann, Johann (1932). Einige Sätze über messbare Abbildungen. *Ann. Math. 33*: 574–586, and *Collected Works* [1961, below], 2, no. 16, p. 297.

——— (1940–1941). *Lectures on invariant measures*. Notes by Paul R. Halmos. Unpublished. Institute for Advanced Study, Princeton.

——— (1961–1963). *Collected Works*. Ed. A. H. Taub. Pergamon Press, London.

Oxtoby, John C., and Stanislaw M. Ulam (1939). On the existence of a measure invariant under a transformation. *Ann. Math.* (Ser. 2) *40*: 560–566.

Paplauskas, A. B. (1971). Egorov, Dimitry Fedorovich. *Dictionary of Scientific Biography, 4*, pp. 287–288.

Ravetz, Jerome R., and Ivor Grattan-Guinness (1972). Fourier, Jean Baptiste Joseph. *Dictionary of Scientific Biography 5*, pp. 93–99.

Riesz, Frigyes [Frédéric] (1909). Sur les opérations fonctionnelles linéaires. *Comptes Rendus Acad. Sci. Paris 149*: 974–977.

*——— (1914). Démonstration nouvelle d'un théorème concernant les opérations. *Ann. Ecole Normale Sup.* (Ser. 3) *31*: 9–14.

——— (1930–1932). Sur l'existence de la dérivée des fonctions monotones et sur quelques problèmes qui s'y rattachent. *Acta Sci. Math. Szeged 5*: 208–221.

——— and Béla Szökefalvi-Nagy (1953). *Functional Analysis*. Ungar, New York (1955). Transl. L. F. Boron from 2d French ed., *Leçons d'analyse fonctionelle*. 5th French ed., Gauthier-Villars, Paris (1968).

Rudin, Walter (1966, 1974, 1987). *Real and Complex Analysis*. 1st, 2d and 3d eds. McGraw-Hill, New York.

——— (1976). *Principles of Mathematical Analysis*. 3d ed. McGraw-Hill, New York.

Saks, Stanisław (1937). *Theory of the Integral*. 2d ed. Monografie Matematyczne, Warsaw; English transl. L. C. Young. Hafner, New York. Repr. Dover, New York (1964).

Schaerf, H. M. (1947). On the continuity of measurable functions in neighborhood spaces. *Portugal. Math. 6*: 33–44.

——— (1948). On the continuity of measurable functions in neighborhood spaces II. *Ibid. 7*: 91–92.

Ulam, Stanislaw Marcin (1976). *Adventures of a Mathematician*. Scribner's, New York.

*Vitali, Giuseppe (1904–05). Sulle funzioni integrali. *Atti Accad. Sci. Torino 40*: 1021–1034.

*——— (1908). Sui gruppi di punti e sulle funzioni di variabili reali. *Ibid. 43*: 75–92.

Zakon, Elias (1965). On "essentially metrizable" spaces and on measurable functions with values in such spaces. *Trans. Amer. Math. Soc. 119*: 443–453.

Zygmund, Antoni (1959). *Trigonometric Series*. 2 vols. 2d ed. Cambridge University Press.



第8章　概率论初步

有限集上定义概率是最容易的．例如，考虑掷一个均匀的硬币．这里“公平”的意思是正面和反面等可能地出现．这种情况可以表示为有两个点 H 和 T 的集合，其中 H =“正面”，T =“反面”．所有可能结果的概率和是等于1的集合．令“$P(\cdots)$”表示“…的概率”．如果两个可能的结果不同时发生，那么可以假设把它们的概率相加．所以有 $P(H)+P(T)=1$．若假设 $P(H)=P(T)$，则 $P(H)=P(T)=1/2$．

现在假设掷币两次，则两次掷币可能有四种结果：HH、HT、TH 和 TT．若这四个事件等可能发生，则它们每个必有概率1/4．同样，如果掷币 n 次，则有 2^n 种可能的字符串，每个字符串由 n 个字母 H 和 T 组成，每一串出现的概率为 $1/2^n$．

然后令 n 趋于无穷，则得到 H 和 T 的所有可能的无穷序列．每一个单独的序列概率为0，但是这并不能像 n 有限时那样确定其他可能结果集合发生的概率．考虑这样的集合，首先用1来代替 H，0来代替 T，在序列之前用“二进制小数点”（就像十进制小数点那样），且把这个序列看成一个二进制展开．例如，

$$0.0101010101\cdots \text{是 } 1/4+1/16+\cdots=1/3 \text{ 的展开，}$$

一般而言，如果对所有的 n，有 $d_n=0$ 或1，那么这个序列或二进制展开 $0.d_1d_2d_3\cdots=\sum_{1\leqslant n<\infty}d_n/2^n$．在区间[0，1]上的每一个数字 x 都有这样一个展开．如果 k 是一个满足 $0<k<2^n$ 的整数，则二进有理数 $k/2^n$ 有两个这样的展开，一个是 $k/2^n$ 的一般展开接着都是0，另一个是 $(k-1)/2^n$ 的一般展开接着一串无穷的1．有可数多个这样的二进有理数．[0，1]中的所有其他数字都有唯一的二进制展开．

在区间[0，1]，有一个通常的测度 λ，称为勒贝格测度或一致测度，它为每一个子区间的长度赋值．那么 λ 是定义在所有博雷尔集的 σ-代数 $\mathcal{B}$ 上，或者是定义在所有勒贝格可测集的较大 σ-代数 $\mathcal{L}$ 上的一个可数可加测度．这些二进有理数的可数集对于 λ 来说测度为0．

250

n 次掷币结果的一个可能序列对应于以给定的 n 个0或1的序列开始的所有无穷序列的集合，依次对应于[0，1]上长度为 $1/2^n$ 的子区间．任意一个由 n 次掷币得出的 m 个不同结果序列的集合与[0，1]中勒贝格测度为 $m/2^n$ 的集合相对应．从这种程度上来说，至少，勒贝格测度对应着前面定义的概率．这是一个说明测度和概率一般关系的例子，测度和概率的关系将在下面定义．

8.1　基本定义

可测空间 $(\Omega,\mathcal{S})$ 是一个具有 Ω 的子集的 σ-代数 $\mathcal{S}$ 的集合 Ω．概率测度（probability measure）P 是 $\mathcal{S}$ 上满足 $p(\Omega)=1$ 的测度．（回忆测度的定义，测度是非负的、可数可加的．）那么 $(\Omega,\mathcal{S},P)$ 称为概率空间（probability space）．

如果 Ω 是有限的，如在有限（甚至是可数）次掷币的情况中，通常 σ-代数 $\mathcal{S}$ 是 2^Ω，即 Ω 的所有子集的集合．如果 Ω 是不可数的，比如 $\Omega=[0,1]$，通常 $\mathcal{S}$ 是 2^Ω 的真子集．$\mathcal{S}$ 的元素在测度论中称为可测集，在概率空间中又称为事件（event）．对一个事件 A，有时 $P(A)$ 可以写成 PA．

在一个有 m 个元素的有限集 F 上，有一个一般概率测度 P，称为 2^F 上的（离散的）一致测度

(uniform measure)，它对 F 上的每一个 x 的单元素集 $\{x\}$ 赋予 $1/m$ 的概率. 在掷币的例子中，$m=2^n$，$n=1, 2, \cdots$. 另一个经典的例子是一个均匀的骰子(die)，即随机地扔一个完美的立方体，其中各面标有从 1 ~ 6 的整数，6 个面中每个面出现的概率都为 1/6.

如果$(\Omega, \mathcal{A}, P)$是一概率空间且$(S, \mathcal{B})$是任意可测空间，一个从 Ω 映射到 S 的可测函数称为随机变量(random variable). 那么定义在$\mathcal{B}$上的像测度 $P\circ X^{-1}$是一个概率测度，称为 X 或$\mathcal{L}(X)$法则.

通常，S 为实轴$\mathbb{R}$ 且$\mathcal{B}$为博雷尔集的 σ-代数. 一个实值随机变量 X 的期望(expectation)或者说是均值(mean)EX 定义为 $\int X\mathrm{d}P$ 当且仅当这个积分存在. 与一般的积分相同，该期望是线性的.

8.1.1 定理 *对任意两个使得期望 EX 和 EY 都有定义且有限的随机变量 X 和 Y，以及任意的常数 c，$E(cX+Y)=cEX+EY$.*

251

证明 这是定理 4.1.10 的一部分. □

随机变量 X 的方差(variance)$\sigma^2(X)$定义为

$$\operatorname{var}(X) := \sigma^2(X) := \begin{cases} E(X-EX)^2 = EX^2-(EX)^2, & 若\ EX^2<\infty \\ +\infty, & 若\ EX^2=+\infty. \end{cases}$$

(这里，“:=”始终意味着“定义中的等于.”)如果 $E(X^2)<\infty$，则 $\sigma(X):=\sqrt{\sigma^2(X)}$ 称为 X 的标准差(standard deviation). 由 Cauchy-Bunyakovsky-Schwarz 不等式(5.1.4)，$E|X|=E(|X|\cdot 1)\leqslant (EX^2)^{1/2}$，所以如果 X 是随机变量且 $E(X^2)<\infty$，则 EX 是有定义的且有限的.

事件经常以$\{\omega: \cdots\omega\cdots\}$的形式给出，即所有满足给定条件的 ω 的集合. 在概率中，一事件经常仅仅写成条件，而没有自变量 ω. 例如，如果 X 和 Y 是两个随机变量，人们往往记作 $P(X>Y)$，而不使用长符号 $P(\{\omega: X(\omega)>Y(\omega)\})$.

在掷币过程中，假设第一掷得到 H，则我们假设第二掷得到 H 或 T 是等可能发生的，令 HH 和 HT 的概率都是 1/4. 换言之，第二次掷币的结果和第一次的结果是独立的. 这是概率论中至关重要的概念的一个例子，后面将给出更一般的定义.

两事件 A 和 B 称为是独立的(independent，对于概率测度 P)，当且仅当 $P(A\cap B)=P(A)P(B)$. 令 X 和 Y 为同一概率空间的两个随机变量，并且分别在 S 和 T 中取值，其中$(S, \mathcal{U})$和$(T, \mathcal{V})$是对 X 和 Y 可测的可测空间. 那么对于 Ω 中的每一个 ω，我们能组成一个“向量”随机变量$\langle X, Y\rangle$，其中$\langle X, Y\rangle(\omega):=\langle X(\omega), Y(\omega)\rangle$. X 和 Y 称为独立的当且仅当$\mathcal{L}(\langle X, Y\rangle)$等于积测度$\mathcal{L}(X)\times\mathcal{L}(Y)$. 换言之，对任意可测集 $U\in\mathcal{U}$和 $V\in\mathcal{V}$，$P(X\in\mathcal{U}, Y\in\mathcal{V})=P(X\in\mathcal{U})P(Y\in\mathcal{V})$. 如果 $S=T=\mathbb{R}$，X 和 Y 是独立的，$E|X|<\infty$，且 $E|Y|<\infty$，则由 Tonelli-Fubini 定理(4.4.5，和定理 4.1.11)得，$E(XY)=EXEY$.

给定任意的概率空间$(\Omega_j, \mathcal{S}_j, P_j)$，$j=1, \cdots, n$，我们能构造笛卡儿积 $\Omega_1\times\Omega_2\times\cdots\times\Omega_n$，其中由定理 4.4.6 给出 S_j 的积 σ-代数和积概率测度 $P_1\times P_2\times\cdots\times P_n$. 在一个概率空间$(\Omega, \mathcal{S}, P)$上的随机变量 $X_1, \cdots, X_n$ 称为独立的，或者更准确地称为联合独立(jointly independent)，如果$\mathcal{L}(\langle X_1, \cdots, X_n\rangle)$等于积测度$\mathcal{L}(X_1)\times\cdots\times\mathcal{L}(X_n)$. 例如，如果 X_i 是实值的，给定$\mathbb{R}$ 的博雷尔σ-代数$\mathcal{B}$上的任意概率测度 $\mu_1, \cdots, \mu_n$，由定理 4.4.6 知，取概率空间$(\mathbb{R}, \mathcal{B}, \mu_i)$的积可得到$\mathbb{R}^n$ 上满足给定的律的积测度 μ，其中坐标 $X_1, \cdots, X_n$ 是独立的.

252

任意随机变量集称为是独立的，当且仅当其中的每一个有限子集是独立的. 随机变量 X_j 称为是两两独立的(pairwise independent)，当且仅当对所有的 $i\neq j$，X_i 和 X_j 是独立的. (注意，独立的所

有定义都是关于某个概率测度 P 的；随机变量对于某个 P 是独立的，但对另一个可能不是.

在概率空间(Ω, P)上，两个随机变量 X 和 Y 有有限方差，X 和 Y 的协方差(covariance)定义为

$$\operatorname{cov}(X,Y) := E((X-EX)(Y-EY)) = E(XY) - EXEY.$$

(这是 $X-EX$ 和 $Y-EY$ 在希尔伯特空间 $L^2(\Omega, P)$上的内积.)所以如果 X 和 Y 是独立的，它们的协方差是0. 独立的一些好处是与协方差和方差的性质有关的.

8.1.2 定理　对概率空间上任意有有限方差的随机变量 $X_1, \cdots, X_n$，令 $S_n := X_1 + \cdots + X_n$. 则

$$\operatorname{var}(S_n) = \sum_{1\leq i\leq n}\operatorname{var}(X_i) + 2\sum_{1\leq i<j\leq n}\operatorname{cov}(X_i, X_j).$$

如果所有的协方差都是0，从而如果 X_i 是独立的或者恰好两两独立，则

$$\operatorname{var}(X_1 + \cdots + X_n) = \operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n).$$

证明　如果用 $X_i - EX_i$ 代替每一个 X_i，则方差和协方差都不改变，同样 $\operatorname{var}(S_n)$也不改变，由于期望的线性性(定理8.1.1). 所以可以假定对所有的 i，$EX_i = 0$，则

$$\operatorname{var}(S_n) = E(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)$$

以及定理的第一个结论得证，则第二个也得证. □

回忆一下，对任意一个事件 A，示性函数 1_A 定义为

$$1_A(x) = \delta_x(A) = \begin{cases}1, & 若\ x\in A\\ 0, & 其他.\end{cases}$$

(概率学家普遍使用"特征函数"来谈及概率测度 P 的傅里叶变换. 例如，在$\mathbb{R}$上，$f(t) = \int e^{ixt}dP(x)$ 是 P 的特征函数.)由事件所组成的集合称为是独立的，当且仅当它们的示性函数是独立的. 253

可以把掷币的情况进行推广. 假设 $A_1, \cdots, A_n$ 是独立事件，都有相同的概率 $P = P(A_j)$，$j = 1, \cdots, n$. 令 $q := 1 - p$. 例如，一个人掷"有偏差的"硬币 n 次，其中正面的概率是 p. 令 A_j 是在第 j 次掷币时出现正面的事件. n 次结果的任意特殊序列，其中 A_j 发生 k 次，另外 $n-k$ 次不发生，由独立性知，概率为 $p^k q^{n-k}$. 因此，A_j 确切发生 k 次的概率是$\binom{n}{k}$个这样概率的和，即 $b(k, n, p) := \binom{n}{k}p^k q^{n-k}$，其中

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad 对任意整数\ 0\leq k\leq n.$$

那么 $b(k, n, p)$称为二项概率(binomial probability). 这还可以描述为"在 n 次独立试验中有 k 次成功的概率，其中每次试验成功的概率为 p". 对这种概率的更广泛的讨论，以及更具体的、组合方面的概率论，可以参考一本经典的书 W. Feller(1968).

习题

1. 在$\{1, \cdots, n\}$上的一致概率定义为 $P\{j\} = 1/n$，$j = 1, \cdots, n$. 对于恒等随机变量 $X(j) \equiv j$，求均值 EX 和方差 $\sigma^2(X)$.
2. 令 X 为一随机变量，它在区间$[1, 4]$有均匀分布 P，即对任意博雷尔集 A，$P(A) = \lambda(A\cap[1, 4])/3$. 求 X 的均值和方差.
3. 令 X 和 Y 为有有限方差的独立的实值随机变量. 令 $Z = aX + bY + c$. 根据 X，Y 和 a，b，c 求 Z 的均值和方差.

4. 求一个有三个随机变量 X，Y，Z 的概率空间，这三个随机变量是两两相互独立的，但并非独立. [提示：取如习题 1 中 $n=4$ 的一个空间，且 X，Y 和 Z 是示性函数.]
5. 求一测度空间$(\Omega, \mathcal{S})$，它有两个$\mathcal{S}$上的概率测度 P 和 Q 以及$\mathcal{S}$中的两个集合 A 和 B，这两个集合相对 P 是独立的，但对 Q 不独立.
6. 一个由平行线集构成的平面，其中平行线间的距离为 d. 一个长度为 s 的针随机地投在平面上. 具体地，令 X 为针的中心到最近线的距离. 令 θ 为在针和确定平面的线之间所成的最小非负角. 那么对于某个 c 和 γ，$P(a\leqslant X\leqslant b)=(b-a)/c$，其中 $0\leqslant a\leqslant b\leqslant c$，$P(\alpha\leqslant\theta\leqslant\beta)=(\beta-\alpha)/\gamma$，其中 $0\leqslant\alpha\leqslant\beta\leqslant\gamma$，且 θ 与 X 独立. 254
   (a)对于一次"随机"投掷，c 和 r 应该是何值?
   (b)求针与一条或多条线相交的概率. [提示：分 $s\geqslant d$ 和 $s<d$ 两种情况.]
7. 假设不是针，而是投掷一个半径为 r 的圆形硬币. 那么它至少与一条线相交的概率是多少?
8. 求在 10 次独立试验中恰有 3 次成功的概率，其中每次成功的概率为 0.3.
9. 设 $E(k, n, p)$ 为 n 次独立试验中有 k 次或更多次成功的概率，其中每次试验成功的概率为 p，所以 $E(k,n,p)=\sum_{k\leqslant j\leqslant n} b(j,n,p)$. 证明：
   (a)对 $k=1, \cdots, n$，$E(k, n, p)=1-E(n-k+1, n, 1-p)$.
   (b)对 $k=1, 2, \cdots$，$E(k, 2k-1, 1/2)=1/2$.
10. 对在同一空间(Ω, P)上的随机变量 X 和 Y，如果 $\mathrm{var}(X)>0$ 且 $\mathrm{var}(Y)>0$，令
$$r(X,Y)=\mathrm{cov}(X,Y)/(\mathrm{var}(X)\mathrm{var}(Y))^{1/2}(\text{"相关系数"});$$
   如果 $\mathrm{var}(X)$ 或 $\mathrm{var}(Y)$是 0 或∞，$r(X, Y)$是无定义的.
   (a)证明：如果 $r(X, Y)$是有定义的，则 $r(X, Y)=\cos\theta$，其中 θ 是由这两个函数生成的 $L^2(P)$的子空间中向量 $X-EX$ 和 $Y-EY$ 所成的角，所以 $-1\leqslant r(X, Y)\leqslant 1$.
   (b)如果 $r(X, Y)=0.7$ 且 $r(Y, Z)=0.8$，求 $r(X, Z)$的最小和最大可能值.

8.2 概率空间的无穷积

对于 n 次重复而独立试验序列，当 n 趋于∞时，某些概率分布和变量是收敛的. 在证明这样的极限定理中，构造一个概率空间是非常有用的，在其上独立的随机变量序列是以自然的方式定义的. 特别地，可定义为一个可数笛卡儿积的坐标.

有限多个 σ-有限测度空间的笛卡儿积给出了一个 σ-有限测度空间(定理 4.4.6). 例如，直线上勒贝格测度的笛卡儿积给出了有限维欧几里得空间的勒贝格测度. 但是，假设取一个有两个点的测度空间$\{0, 1\}$，其中每一个点的测度为 1，即 $\mu(\{0\})=\mu(\{1\})=1$，并且构成了这个空间的一个可数笛卡儿积，所以集合的任意可数积的测度等于它们测度的积. 然后取所有测度为 1 的单元素集构成的不可数空间，给定的测度通常称为"可数测度". 一个具有可数测度的不可数集并非是 σ-有限测度空间，尽管在这个例子中它是有限测度空间的可数积. 与此相反，概率空间的可数积仍然是一个概率空间. 以下是一些定义. 255

对每一个 $n=1, 2, \cdots$，设$(\Omega_n, \mathcal{S}_n, P_n)$为一个概率空间. Ω 为笛卡儿积 $\Pi_{1\leqslant n<\infty}\Omega_n$，即对所有的 n，所有满足 $\omega_n\in\Omega_n$ 的序列 $\{\omega_n\}_{1\leqslant n<\infty}$ 的集合. 对每一个 n，令 π_n 为 Ω 到 Ω_n 的自然投影. $\pi_n(\{\omega_m\}_{m\geqslant 1})=\omega_n$. $\mathcal{S}$为 Ω 子集的最小 σ-代数，使得对所有的 m，π_m 从$(\Omega, \mathcal{S})$到$(\Omega_m, \mathcal{S}_m)$是可测的. 换言之，对所有的 n，$\mathcal{S}$是包含了所有集合 $\pi_n^{-1}(A)$和 $A\in\mathcal{S}_n$的最小 σ-代数.

例如，区间$[0, 1]$与通过十进制或二进制展开的一个积是相关的：对每一个 $n=1, 2, \cdots$令 Ω_n 为 10 个数字$\{0, 1, \cdots, 9\}$的集合，其中 $P_n(\{j\})=1/10$，$j=0, 1, \cdots, 9$. 那么 Ω 中的点

$\omega=\{j_n\}_{n\geq 1}$产生了数字$x(\omega)=\sum_{n\geq 1} j_n/10^n$，其中$j_n$是$x$的十进制展开中的数字．所有正整数的集合都能写成无限多个可数无限集A_k的一个并集(例如，令A_k为对于2^{k-1}可分，但对于2^k不可分的整数集合，其中$k=1，2，\cdots$)．令$A_k:=\{n(k,i)\}_{i\geq 1}$，则在$[0,1]$中的每一个数字$x$(有十进制展开$\sum_{n\geq 1} j_n(x)/10^n$)能够用数字$y_k$编码一个$y_k:=\sum_i j_{n(k,i)}/10^i$的无穷序列．令$T(x):=\{y_k\}_{k\geq 1}$且$T_k(x):=y_k$．设$\lambda$为$[0,1]$上的勒贝格测度，则$\lambda\circ T^{-1}$给出了$[0,1]$中序列积上的一个概率测度，其中任意$(y_1,\cdots,y_k)$的分布是在$\mathbb{R}^k$中单位立方体$I^k$上的勒贝格测度$\lambda^k$．如果空间$\Omega_n$上的概率测度$P_n$对于某个从$[0,1]$映射到$\Omega_n$的函数能够表示成$P_n=\lambda\circ V_n^{-1}$，同样对许多的$P_n$也成立，那么测度$P_n$的积可记作$\lambda\circ W^{-1}$，其中$W(x):=\{V_n(T_n(x))\}_{n\geq 1}$．

另一方面，有些概率测度不能写成一维勒贝格测度的像．因为积空间上的测度不同于积测度，在商空间Ω_n的任意有限笛卡儿积上的测度的相容定义并不能保证存在无穷积上的一个相容测度(12.1节习题2)．所以本节对完全任意的概率空间序列给出笛卡儿积测度是没有价值的．用传统的方法(比如十进制展开)不能证明这个积是存在的(至少不是直接的)．注意到这个一般性，下面回到构造上来．

令$\mathcal{R}$为所有集合$\prod_n A_n\subset\Omega$组成的集族，其中对所有的n，$A_n\in\mathcal{S}_n$，且除去至多有限多个n的值，有$A_n=\Omega_n$．$\mathcal{R}$的元素称为矩形或有限维矩形．现在回顾一下3.2节半环的概念．$\mathcal{R}$有下面的性质．

256

8.2.1 命题 *无穷积Ω上的有限维矩形的集族$\mathcal{R}$是半环．由$\mathcal{R}$产生的代数$\mathcal{A}$是$\mathcal{R}$中元素的有限不交并的集族．*

证明 如果C和D是任意两个矩形，则显然$C\cap D$是一个矩形．在两个空间的积中，由命题3.2.2知，矩形的集族是一个半环．特别地，两个矩形的差是两个不交矩形的并：

$$(A\times B)\setminus(E\times F)=((A\setminus E)\times B)\cup((A\cap E)\times(B\setminus F)).$$

由归纳可得，在任意有限的笛卡儿积中，两个矩形的任意差$C\setminus D$是矩形的一个有限不交并．从而$\mathcal{R}$是一个半环．我们有$\Omega\in\mathcal{R}$，因此由$\mathcal{R}$产生环$\mathcal{A}$，如同在命题3.2.3中，环$\mathcal{A}$是一个代数．因为每一个代数是一个环，$\mathcal{A}$是$\mathcal{R}$产生的代数．由命题3.2.3知，$\mathcal{A}$包含了所有$\mathcal{R}$中元素的有限不交并． □

现在对$A=\prod_n A_n\in\mathcal{R}$，令$P(A):=\prod_n P_n(A_n)$．因为几乎有限多个因子都是1，积是收敛的．下面是在本节的剩余部分将要证明的主要定理．

8.2.2 定理(无穷积概率的存在性定理) *$\mathcal{R}$上的P可唯一地扩张成$\mathcal{S}$上的一个(可数可加的)概率测度．*

证明 对每一个$A\in\mathcal{A}$，把A记作$\mathcal{R}$中集合的一个有限不交并(命题8.2.1)，表示成$A=\bigcup_{1\leq r\leq N} P(B_r)$，且定义$P(A):=\sum_{1\leq r\leq N} P(B_r)$．首先证明$P$在有限不交的并中是明确定义的且有限可加的．对所有的$n\geq n(r)$，某个$n(r)<\infty$，每一个B_r是集$A_{rn}\in\mathcal{S}_n$的积，且$A_{rn}=\Omega_n$．对$r=1,\cdots,N$，令m为$n(r)$的最大值．因为对$n>m$，所有的A_{rn}等于Ω_n．这些集合上的P的性质与$\Omega_1\times\cdots\times\Omega_m$上的有限积测度的性质等价．为证明$P$是明确定义的，如果一个集合等于$\mathcal{R}$中集合的两个不同的有限不交并，对这两个并可以取$m$的最大值，且仍有一个有限积．由有限积测度定理(4.4.6)，P是明确定义的，且在$\mathcal{A}$上有限可加．

如果P在$\mathcal{A}$上可数可加，那么由定理3.1.10知，它在$\mathcal{S}$上有一个唯一的可数可加扩张．所以只

[257] 需证明$\mathcal{A}$上的可数可加性．等价地，由定理 3.1.1 知，如果 $A_j \in \mathcal{A}$，$A_1 \supset A_2 \supset \cdots$，且 $\bigcap_j A_j = \varnothing$，我们证明 $P(A_j) \downarrow 0$．换句话说，如果 A_j 是$\mathcal{A}$中递减的集合序列，且对某个 $\varepsilon > 0$，对所有的 j，$P(A_j) \geqslant \varepsilon$，需证明 $\bigcap_j A_j \neq \varnothing$．

在$\mathcal{A}$上，设 $P^{(0)} := P$．对每一个 $n \geqslant 1$，令 $\Omega^{(n)} := \Pi_{m>n}\Omega_m$．类似于$\mathcal{A}$和 P 在 Ω 上的定义，在 $\Omega^{(n)}$ 上定义$\mathcal{A}^{(n)}$和 $P^{(n)}$．对每一个 $E \subset \Omega$，$x_i \in \Omega_i$，$i = 1, \cdots, n$，令

$$E^{(n)}(x_1,\cdots,x_n) := \{\{x_m\}_{m>n} \in \Omega^{(n)} : x = \{x_i\}_{i\geqslant 1} \in E\}.$$

对积空间 $X \times Y$ 上的集合 A 和 $x \in X$，设 $A_x := \{y \in Y : \langle x, y\rangle \in A\}$（见图 8-1A）．如果 A 包含在积 σ-代数 $\mathcal{S}\otimes\mathcal{T}$中，由定理 4.4.3 的证明得 $A_x \in \mathcal{T}$．对任意的 $E \in \mathcal{A}$，存在足够大的 n，使得对某个 $F \subset \Pi_{n\leqslant N}\Omega_n$，$E = F \times \Pi_{n>N}\Omega_n$．（因为 E 是具有这个性质的矩形的有限并，对矩形取 N 的最大值．）则 $F = \bigcup_{k=1}^{m} F_k$，其中对某个 $F_{ki} \in \mathcal{S}_i$，$i = 1, \cdots, N$，$k = 1, \cdots, m$，$F_k = \Pi_{i=1}^N F_{ki}$．现在对任意 $n < N$ 和 $x_i \in \Omega_i (i = 1, \cdots, n)$，$E^{(n)}(x_1, \cdots, x_n) = G \times \Omega^{(N)}$，其中对所有的 $i = 1, \cdots, n$，G 是这些集合 $\prod_{n<i\leqslant N} F_{ki}$的并，$x_i \in F_{ki}$．因此，$E^{(n)}(x_1, \cdots, x_n) \in \mathcal{A}^{(n)}$，所以 $P^{(n)}$ 是有定义的．在 $\Omega_1 \times \cdots \times \Omega_n \times \prod_{n<i\leqslant N}\Omega_i$ 应用 Tonelli-Fubini 定理（定理 4.4.6），有

$$P(E) = \int P^{(n)}(E^{(n)}(x_1,\cdots,x_n))\Pi_{1\leqslant j\leqslant n}\mathrm{d}P_j(x_j).$$

对所有的 j，对满足 $P(A_j) \geqslant \varepsilon$ 的 ε，令

$$F_j := \{x_1 \in \Omega_1 : P^{(1)}(A_j^{(1)}(x_1)) > \varepsilon/2\}.$$

对每一个 j，应用(8.2.3)于 $E = A_j$，$n = 1$，则

$$\begin{aligned}\varepsilon \leqslant P(A_j) &= \int P^{(1)}(A_j^{(1)}(x_1))\mathrm{d}P_1(x_1)\\ &= \left(\int_{F_j} + \int_{\Omega_1\setminus F_j}\right)P^{(1)}(A_j^{(1)}(x_1))\mathrm{d}P_1(x_1)\\ &\leqslant P_1(F_j) + \varepsilon/2\end{aligned}$$

[258] （见图 8-2B）．因此对所有的 j，$P_1(F_j) \geqslant \varepsilon/2$．当 j 递增时，集合 A_j 递减．所以，对 $A_j^{(1)}$ 和 F_j 也成立．

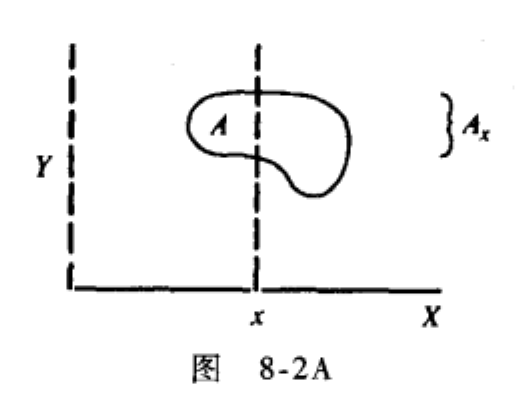


图 8-2A

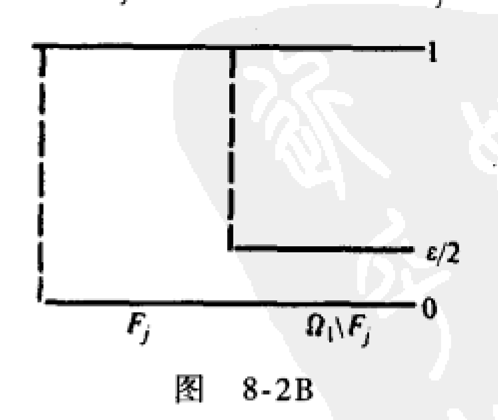


图 8-2B

因为 P_1 是可数可加的，$P_1\left(\bigcap_j F_j\right) \geqslant \varepsilon/2 > 0$（由示性函数的单调收敛性），所以 $\bigcap_j F_j \neq \varnothing$．取任意的 $y_1 \in \bigcap_j F_j$．令 $f_j(y, x) := P^{(2)}(A_j^{(2)}(y, x))$ 且 $G_j := \{x_2 \in \Omega_2 : f_j(y_1, x_2) > \varepsilon/4\}$．那么当 j 增加时 G_j 减小，对所有的 j，有

$$\varepsilon/2 < P^{(1)}(A_j^{(1)}(y_1)) = \int f_j(y_1, x)\, dP_2(x)$$

且 $P_2(G_j) > \varepsilon/4$，故所有 G_j 的交集在 Ω_2 中是非空的，且可选出其中的 y_2.

由归纳，用同样的讨论得，对所有的 n，存在 $y_n \in \Omega_n$，使得对所有的 j，n，$P^{(n)}(A_j^{(n)}(y_1, \cdots, y_n)) \geqslant \varepsilon/2^n$. 令 $y := \{y_n\}_{n\geqslant 1} \in \Omega$. 对每一个 j，为证明 $y \in A_j$，选取足够大的 n（依赖于 j），使得对所有的 $x_1, \cdots, x_n$，$A_j^{(n)}(x_1, \cdots, x_n) = \varnothing$ 或 $\Omega^{(n)}$. 这是可以做到的，因为 $A_j \in \mathcal{A}$. 那么 $A_j^{(n)}(y_1, \cdots, y_n) = \Omega^{(n)}$，因此 $y \in A_j$. 故 $\bigcap_j A_j \neq \varnothing$. □

事实上，定理 8.2.2 对任意概率空间的积（不必是可数的）成立. 证明不需要大的改变，因为 σ-代数$\mathcal{S}$中的每一个集合仅依赖于可数多个坐标. 换句话说，给定一个积 $\prod_{i\in I}\Omega_i$，其中 I 可能是一个不可数的下标集，对$\mathcal{S}$中的每一个集合 A，存在 I 的一个可数子集 J 和一个集合 $B \subset \prod_{i\in J}\Omega_i$，使得 $A = B \times \prod_{i\notin J}\Omega_i$.

习题

1. 在$\mathbb{R}^2$ 中，对由区间的笛卡儿积构成的普通矩形（在每一个端点处或许开或许闭），证明：对任意两个矩形 C 和 D，对某个有限的 k，$C \setminus D$ 是至多 k 个矩形的并，再求 k 的最小可能值.
2. 类似于定理 3.2.7，对任意一个 σ-代数$\mathcal{B}$，如果$\mathcal{B}$上的任意两个概率测度，在$\mathcal{C}$上是相等的，在$\mathcal{B}$上也相等，那么集族$\mathcal{C}\subset\mathcal{B}$称为一个概率确定类（probability determining class）. 回顾对于$\mathcal{C}$产生了$\mathcal{B}$并不需要$\mathcal{C}$是一个（概率）确定类.
   (a) 证明：在一个空间$(\Omega_n, \mathcal{S}_n)$的可数积$(\Omega, \mathcal{S})$中，所有矩形的集合对$\mathcal{S}$是一个概率确定类.
   (b) 设$\mathbb{R}_m$ 为所有矩形 $\bigcap_{j\in F}\pi_j^{-1}(A_j)$ $(A_j \in \mathcal{S}_j)$ 的集合，其中 F 包含 m 个下标. 证明：对于任意有限的 m，$\mathbb{R}_m$ 不是一个概率确定类，例如，如果每一个 Ω_n 是一个两点的集合$\{0, 1\}$且$\mathcal{S}_n$是所有它的子集的 σ-代数. [提示：令 $\omega_2, \omega_3, \cdots$ 是独立的且每一个等于 0 或 1，概率为 1/2. 令 $\omega_1 = 0$ 或 1，其中 $\omega_1 \equiv S_m := \omega_2 + \cdots + \omega_{m+1} \bmod 2$（即 $\omega_1 - S_m$ 可被 2 整除）. 证明当给定的法则对所有的 ω_j 独立时，$\mathbb{R}_m$ 中的每一个集合都有相同的概率.]

259

3. 证明：对概率空间的任意（不可数）积，定理 8.2.2 成立，正如本节的末尾所提出的那样. [提示：证明所描述的所有集合 A 的集族是一个 σ-代数，利用集合 J_m 的一可数并是可数的这一事实.]
4. 设$(X, \mathcal{S}, P)$是一个概率空间且$(Y, \mathcal{T})$是一个可测空间. 假设对每一个 $x \in X$，Q_x 是$(Y, \mathcal{T})$上的概率测度. 假设对每一个 $C \in \mathcal{T}$，函数 $x \mapsto Q_x(C)$ 对$\mathcal{S}$是可测的. 证明：有一个$(X\times Y, \mathcal{S}\otimes\mathcal{T})$上的概率测度 μ，使得对任意集合 $B \in \mathcal{S}\otimes\mathcal{T}$，
$$\mu(B) = \iint 1_B(x, y)\, dQ_x(y)\, dP(x).$$
5. 假设对 $n = 1, 2, \cdots$ 且 $0 \leqslant t \leqslant 1$，$P_{n,t}$是$\mathbb{R}$ 中博雷尔 σ-代数上的一个概率测度，使得对区间$[a, b]\subset\mathbb{R}$ 和每一个 n，如果$f(t) := P_{n,t}([a, b])$，则 f 是博雷尔可测的. 令 P_t 为在$\mathbb{R}^\infty$ 上的积测度 $\Pi_{1\leqslant n<\infty}P_{n,t}$. 证明：存在$\mathbb{R}^\infty$ 上的概率测度 Q，使得对每一个可测集 $C \subset \mathbb{R}^\infty$，$Q(C) = \int_0^1 P_t(C)\, dt$.
6. 设 $X := \{0, 1\}$，$\mathcal{S} = 2^X$ 且 $P(\{0\}) = P(\{1\}) = 1/2$. $(\Omega, \mathcal{B}, Q)$为$(X, \mathcal{S}, P)$的一个可数积. 对每一个$\{x_n\}_{n\geqslant 1} \in \Omega$，其中对所有的 n，$x_n = 0$ 或 1，令 $f(\{x_n\}) = \sum_n x_n/2^n$. 证明：像测度$Q\circ f^{-1}$是$[0, 1]$上的勒贝格测度 λ.
7. 对 $j = 1, 2, \cdots$，设$(X_j, \mathcal{S}_j)$为可测空间，P_1 为 $\mathcal{S}_1$上的概率测度. 假设对每一个 n 和每一个

$x_j \in X_j$ $(j=1, 2, \cdots, n)$，$P(x_1, \cdots, x_n)(\cdot)$是$(X_{n+1}, \mathcal{S}_{n+1})$上的概率测度，满足如下性质：对每一个$A \in \mathcal{S}_{n+1}$，$(x_1, \cdots, x_n) \mapsto P(x_1, \cdots, x_n)(A)$对于积$\sigma$-代数$\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$是可测的. 证明：在积$\sigma$-代数$\prod_{n\geq 1} X_n$中存在一个概率测度$P$，使得对每一个$n$，$X_1 \times \cdots \times X_n$中的积$\sigma$-代数上的每一个集合$B$和$C := B \times X_{n+1} \times X_{n+2} \times \cdots$，$P(C) = \int \cdots \int 1_B(x_1, \cdots, x_n) \mathrm{d}P(x_1, \cdots, x_{n-1})(x_n) \cdots \mathrm{d}P(x_1)(x_2) \mathrm{d}P_1(x_1)$. [提示：首先证明这个结论对有限积成立($n$固定). $n=2$的情况即习题4. 由归纳法，证明它到一般有限的n，记$X_1 \times X_2 \times X_3$为$X_1 \times (X_2 \times X_3)$，依此类推. 像在定理8.2.2的证明中从有限积推到无穷积那样进行.]

8.3 大数定律

设$(\Omega, \mathcal{S}, P)$为一概率空间. 令$X_1, X_2, \cdots$为Ω上的实值随机变量. 令$S_n := X_1 + \cdots + X_n$. 任意概率为1的事件称为几乎必然(almost surely, a. s.)发生. 因此，(实)随机变量序列Y_n称为几乎必然收敛到随机变量Y，当且仅当$P(Y_n \to Y) = 1$. (随机变量变量序列收敛的集合是可测的，在定理4.2.5中已证). 序列Y_n称为依概率收敛(Converge in probability)到Y，当且仅当对$\forall \varepsilon > 0$，$\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0$. 显然，几乎必然收敛意味着依概率收敛. 序列$X_1, X_2 \cdots$称为满足强大数定律(strong law of large numbers)，当且仅当对某个常数c，S_n/n几乎必然收敛到一个常数c. 序列称为满足弱大数定律(weak law of large numbers)，当且仅当对某个常数c，S_n/n依概率收敛到常数c.

大数定律的一个基本例子是相对频率收敛到概率. 假设X_j是满足对所有的j有$P(X_j = 1) = p = 1 - P(X_j = 0)$的独立变量. 如果$X_j = 1$，就说第$j$次试验成功，否则失败. 那么$S_n$是成功的次数，$S_n/n$是在前$n$次试验中成功的相对频率. 大数定律指出成功的相对频率收敛到成功的概率P.

作为介绍，弱大数定律在有用的但不是最弱的假定下将很快得到证明. 首先来关注一个经典的且经常用到的事实.

8.3.1 Bienaymé-Chebyshev 不等式 对任意的实随机变量X和$t > 0$，

$$P(|X| \geq t) \leq EX^2/t^2.$$

证明 $EX^2 \geq E(X^2 1_{\{|X| \geq t\}}) \geq t^2 P(|X| \geq t)$. □

8.3.2 定理 如果$X_1, X_2, \cdots$是均值为0的随机变量，$EX_j^2 = 1$，对所有的$i \neq j$，$EX_iX_j = 0$，则弱大数定律对它们成立.

注： 假设条件说明X_i在希尔伯特空间$L^2(\Omega, P)$是正交的.

证明 利用定理8.1.2，有$ES_n^2 = n$. 从而$E((S_n/n)^2) = 1/n$，所以对$\forall \varepsilon > 0$，由切比雪夫(Chebyshev)不等式，

$$P(|S_n/n| \geq \varepsilon) \leq 1/(n\varepsilon^2) \to 0, \quad 当\ n \to \infty\ 时.$$ □

随机变量X_j称为同分布的(identically distributed)，当且仅当对所有的n，有$\mathcal{L}(X_n) = \mathcal{L}(X_1)$. "独立同分布"可缩写成"i. i. d."例如，如果$X_1, X_2, \cdots$是独立同分布随机变量且均值为μ，方差为$\sigma^2 (\sigma > 0)$，那么对变量$(X_j - \mu)/\sigma$应用定理8.3.2得，S_n/n依概率收敛到μ. 在大数定律中，对于独立同分布变量，$(X_1 + \cdots + X_n)/n \to c$，通常$c = EX_1$，故$X_1, \cdots, X_n$的均值收敛到"真正的"均值$EX_1$.

在处理独立和几乎必然收敛时，下面的两个事实是很有用的.

8.3.3 引理 如果对所有的n，$0 \leq p_n < 1$，那么$\prod_n (1 - p_n) = 0$当且仅当$\sum_n p_n = +\infty$.

证明 如果$\limsup_{n\to\infty} p_n > 0$，那么显然$\sum_n p_n = +\infty$且$\prod_n (1 - p_n) = 0$. 所以假设当$n \to \infty$时，

260
261

$p_n \to 0$. 因为对 $0 \leqslant p \leqslant 1$, $1-p \leqslant e^{-p}$(在 0 处取等号，且求导数为 $-1 < -e^{-p}$), $\sum p_n = +\infty$ 意味着 $\prod_n (1-p_n) = 0$. 相反地，对 $0 \leqslant p \leqslant 1/2$, $1-p \geqslant e^{-2p}$(在 $p=0$ 和 $1/2$ 时不等式成立，且函数 $f(p) := 1-p-e^{-2p}$ 有一个导数 $f'(p)$ 恰好在 $p = (\ln 2)/2$ 处为 0，是 f 的一个相对极大值). 取 M 足够大，使得对 $n \geqslant M$ 有 $p_n < 1/2$，如果 $\prod_n (1-p_n) = 0$，则 $0 = \prod_{n>M}(1-p_n) \geqslant \prod_{n>M} \exp(-2p_n) = \exp(-2\sum_{n>M} p_n) \geqslant 0$，所以 $\sum_n p_n = +\infty$. □

定义 给定一个概率空间$(\Omega, \mathcal{S}, P)$和一事件序列 A_n，令 $\limsup A_n$ 为事件 $\bigcap_{m \geqslant 1} \bigcup_{n \geqslant m} A_n$.

事件 $\limsup A_n$ 有时被称为"A_n i. o."，意思是当 $n \to \infty$ 时，A_n"无穷多次"出现. 事件 $\limsup A_n$ 出现当且仅当无穷多个 A_n 出现. 例如，令 $Y_n (n=0, 1, \cdots)$ 为随机变量，且对每一个 $\varepsilon > 0$，令 $A_n(\varepsilon)$ 为事件 $\{|Y_n - Y_0| > \varepsilon\}$. 那么 $Y_n \to Y_0$ a. s. 当且仅当对任意 $\varepsilon > 0$ 或等价地对任意 $\varepsilon = 1/m (m = 1, 2, \cdots)$，$A_n(\varepsilon)$不是无穷多次出现.

下面是在概率论中经常用到的事实之一.

8.3.4 定理(博雷尔–坎泰利引理) 如果 A_n 是满足 $\sum_n P(A_n) < \infty$ 的任意事件，那么 $P(\limsup A_n) = 0$. 如果 A_n 是独立的且 $\sum_n P(A_n) = +\infty$，那么 $P(\limsup A_n) = 1$.

证明 第一部分成立，因为对每一个 m，当 $m \to \infty$ 时，

$$P(\limsup A_n) \leqslant P\left(\bigcup_{n \geqslant m} A_n\right) \leqslant \sum_{n \geqslant m} P(A_n) \to 0,$$

262

其中由引理 3.1.5 和定理 3.1.10 得，$P\left(\bigcup_n A_n\right) \leqslant \sum_n P(A_n)$("布尔不等式").

如果 A_n 是独立的且 $\sum_n P(A_n) = +\infty$，则对每一个 m，应用引理 8.3.3 得，

$$P\left(\Omega \setminus \bigcup_{n \geqslant m} A_n\right) = \prod_{n \geqslant m} (1 - P(A_n)) = 0$$

因此对所有的 m，$P\left(\bigcup_{n \geqslant m} A_n\right) = 1$. 令 $m \to \infty$，定理得证. □

一个均匀的硬币掷 n 次，反面每次出现的概率是 $1/2^n$. 因为这些概率的总和是收敛的，所以正面是几乎必然出现的.

下面的定理是关于独立同分布随机变量的强大数定律，是本节的主要定理. 它表明对于强大数定律的成立，$E|X_1| < \infty$ 是充分必要的.（这对于弱大数定律不是十分必要. 本节的注释将会对此进行详细说明.）

8.3.5 定理 对于独立同分布实值 X_j，如果 $E|X_1| < \infty$，那么强大数定律成立，即 $S_n/n \to EX_1$ a. s. 如果 $E|X_1| = +\infty$，则 S_n/n 几乎必然不收敛到任意有限极限.

证明 首先会用到一个引理.

8.3.6 引理 对于任意的非负随机变量 Y，

$$EY \leqslant \sum_{n \geqslant 0} P(Y > n) \leqslant EY + 1.$$

因此，$EY < \infty$ 当且仅当 $\sum_{n \geqslant 0} P(Y > n) < \infty$.

证明 令 $A(k) := A_k := \{k < Y \leqslant k+1\} (k = 0, 1, \cdots)$，那么

$$\sum_{n\geq 0} P(Y > n) = \sum_{n\geq 0}\sum_{k\geq n} P(A_k) = \sum_{k\geq 0}(k+1)P(A_k),$$

由引理 3.1.2，重排非负项的和. 令 $U := \sum_{k\geq 0} k1_{A(k)}$，则 $U\leqslant Y\leqslant U+1$，因此 $EU\leqslant EY\leqslant EU+1\leqslant EY+1$，则引理得证. □

现在继续证明定理 8.3.5，首先假设 $E|X_1|<\infty$. 注意，对任意独立随机变量 X_j 和博雷尔可

[263] 测函数 f_j，变量 $f_j(X_j)$ 也是相互独立的. 特别地，正部 $X_j^+ := \max(X_j, 0)$ 是相互独立的. 类似地，$X_j^- := -\min(X_j, 0)$ 也是相互独立的. 因此，只需证明正部和负部分别收敛. 故我们可以假设对所有的 n，$X_n\geqslant 0$.

令 $Y_j := X_j 1\{X_j\leqslant j\}$，其中 $1\{\cdots\}$ 是示性函数 $1_{\{\cdots\}}$. 令 $T_n := \sum_{1\leqslant j\leqslant n} Y_j$. 给定任意数 $\alpha>1$，令 $k(n) := [\alpha^n]$，其中 $[x]$ 表示小于等于 x 的最大整数，则

$$1\leqslant k(n)\leqslant \alpha^n < k(n)+1\leqslant 2k(n),$$

所以 $k(n)^{-2}\leqslant 4\alpha^{-2n}$. 对 $x\geqslant 1$，$[x]\geqslant x/2$. 取 $\forall\varepsilon>0$. 回顾 $\mathrm{var}(X) := E((X-EX)^2)$ 表示一个随机变量 X 的方差. 由切比雪夫不等式 (8.3.1) 和定理 8.1.2，存在仅依赖于 ε 和 α 的有限常数 $c_1, c_2, \cdots$，使得

$$\begin{aligned}\sum := \sum_{n\geq 1} P\{|T_{k(n)} - ET_{k(n)}| > \varepsilon k(n)\} &\leqslant c_1\sum_{n\geq 1}\mathrm{var}(T_{k(n)})/k(n)^2 \\ = c_1\sum_{n\geq 1} k(n)^{-2}\sum_{1\leqslant i\leqslant k(n)}\mathrm{var}(Y_i) &= c_1\sum_{i\geq 1}\mathrm{var}(Y_i)\sum_{k(n)\geq i} k(n)^{-2},\end{aligned}$$

且

$$\sum_n k(n)^{-2}1_{k(n)\geq i}\leqslant 4\sum_n \alpha^{-2n}1\{\alpha^n\geqslant i\}\leqslant 4i^{-2}/(1-\alpha^{-2})\leqslant c_2 i^{-2},$$

所以如果 F 是 X_1 的法则，$F(X)\equiv P(X_1\leqslant x)$，

$$\begin{aligned}\sum &\leqslant c_3\sum_{i\geq 1} EY_i^2/i^2 = c_3\sum_{i\geq 1}\int_0^i x^2\mathrm{d}F(x)/i^2 \\ &= c_3\sum_{i\geq 1} i^{-2}\Big(\sum_{0\leqslant k<i}\int_k^{k+1} x^2\mathrm{d}F(x)\Big)\leqslant c_4\sum_{k\geq 0}\int_k^{k+1} x^2\mathrm{d}F(x)/(k+1),\end{aligned}$$

因为 $Q_k := \sum_{i>k} i^{-2} < \int_k^\infty x^{-2}\mathrm{d}x = 1/k$，如果 $k\geqslant 1$，有 $Q_k\leqslant 2/(k+1)$，而当 $k=0$ 时，有 $Q_0 = 1+Q_1 < 2 = 2/(k+1)$. 因此

$$\sum\leqslant c_4\sum_{k\geq 0}\int_k^{k+1} x\mathrm{d}F(x) = c_4 EX_1 < \infty.$$

所以由博雷尔–坎泰利 (Borel-Cantelli) 引理 (8.3.4)，当 $n\to\infty$ 时，$|T_{k(n)} - ET_{k(n)}|/k(n)\to 0$ a. s.，$EY_n\uparrow EX_1$. 从而 $ET_{k(n)}/k(n)\uparrow EX_1$. 故 $T_{k(n)}/k(n)\to EX_1$ a. s. 由引理 8.3.6 得，

$$\sum_j P(X_j\neq Y_j) = \sum_j P(X_j > j) < \infty,$$

[264] 所以对足够大的 j，如 $j>m(\omega)$，a. s. $X_j = Y_j$. 当 $n\to\infty$ 时，$S_{m(\omega)}/k(n)\to 0$ 且 $T_{m(\omega)}/k(n)\to 0$，所以 $S_{k(n)}/k(n)\to EX_1$ a. s.，当 $n\to\infty$ 时，$k(n+1)/k(n)\to\alpha$，所以对足够大的 n，$1\leqslant k(n+1)/k(n)<\alpha^2$. 那么对 $k(n)<j\leqslant k(n+1)$，

$$S_{k(n)}/k(n)\leqslant \alpha^2 S_j/j\leqslant \alpha^4 S_{k(n+1)}/k(n+1),$$

因此 a. s.，

$$\alpha^{-2}EX_1 \leqslant \liminf_{j\to\infty} S_j/j \leqslant \limsup_{j\to\infty} S_j/j \leqslant \alpha^2 EX_1.$$

令 $\alpha \downarrow 1$，在 $EX_1 < \infty$ 时定理得证.

相反地，如果 S_n/n 收敛到一个有限极限，则显然 $S_n/(n+1)$ 与 S_{n-1}/n 收敛到相同的极限，因此 $X_n/n = (S_n - S_{n-1})/n \to 0$. 但是如果 $E|X_1| = +\infty$，则由引理 8.3.6 和博雷尔–坎泰利引理 (8.3.4)，对于无穷多的 n，a. s. $|X_n| > n$. 因此，$S_n(\omega)/n$ 是柯西序列的概率为 0. □

在博雷尔–坎泰利引理 (8.3.4) 的一半的证明中不需要独立性，$\sum P(A_n) < \infty$ 意味着 $P(A_n \text{ i. o.}) = 0$，利用次可加性 $P\left(\bigcup_n A_n\right) \leqslant \sum_n P(A_n)$. 这里是 $P\left(\bigcup_{1\leqslant i\leqslant n} A_i\right)$ 的下界，也不需要独立性.

8.3.7 定理（Bonferroni 不等式） 对任意事件 $A_i := A(i)$，

$$P\left(\bigcup_{i=1}^n A_i\right) \geqslant \sum_{i=1}^n P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_i \cap A_j).$$

证明 让我们证明对于示性函数有

$$\sum_{i=1}^n 1_{A_i} \leqslant 1_{\bigcup_{(1\leqslant i\leqslant n)} A_i} + \sum_{1\leqslant i<j\leqslant n} 1_{A_i\cap A_j}.$$

一给定的属于 A_i 的点 ω，对于 i 有 $k(k=0, 1, \cdots, n)$ 个值. 如果 $k=0(0\leqslant 0)$，如果 $k=1(1\leqslant 1)$，且如果 $k\geqslant 2$：$k-1\leqslant k(k-1)/2$，示性函数的不等式成立. 那么对两边关于 P 积分就完成了证明.

□

当后面的和是小的时，Bonferroni 不等式是非常有用的，使得对于并的概率给定的下界接近于上界 $\sum_i P(A_i)$. 例如，假设对所有的 i，$p_i := P(A_i) < \varepsilon$ 且 A_i 是两两独立的，即只要 $i \neq j$，A_i 与 A_j 是独立的，则

$$2\sum_{1\leqslant i<j\leqslant n} P(A_i \cap A_j) \leqslant n(n-1)\varepsilon^2,$$

265

其中如果 $n\varepsilon$ 是小的，$n^2\varepsilon^2$ 就更小. 如果 A_i 确实独立，则它们的并的概率是 $1 - \prod_i (1-p_i)$. 如果积是展开的，在 p_i 中的线性和二次项对应着 Bonferroni 不等式的两个和.

除了下面的习题外，在 9.7 节的末尾还有一些关于进一步深入探讨大数定律的其他习题.

习题

1. 设 $A(n)$ 为一独立事件序列且 $p_n := P(A(n))$. 在 p_n 满足什么条件下有 $1_{A(n)} \to 0$.
   (a) 依概率?
   (b) 几乎必然? [提示：应用博雷尔–坎泰利引理.]
2. 证明：对任意实随机变量 X，$p>0$，$t>0$，$P(|X|\geqslant t) \leqslant E|X|^p/t^p$.
3. 如果 $X_1, X_2, \cdots$ 是随机变量且均值为 0，$EX_j^2 = 1$，且对所有的 $i\neq j$，$EX_iX_j = 0$，证明：对于 $\forall \alpha > 1$，$S_n/n^\alpha \to 0$ a. s..
4. 如果 $X_1, X_2, \cdots$ 是随机变量且对所有的 $i \neq j$，有 $EX_iX_j = 0$，$\sup_j EX_j^2 < \infty$，证明：对于 $\forall \alpha > 1/2$，S_n/n^α 依概率收敛到 0.
5. 设 Y 为一标准的指数分布，即对所有的 $t\geqslant 0$，$P(Y>t) = \mathrm{e}^{-t}$. 求 EY 和 $\sum_{n\geqslant 0} P(Y>n)$ 的值，在此情况下检验引理 8.3.6 中的不等式.

6. 证明：对任意的 $c_n>0$，$c_n \to 0$（不考虑速度），存在随机变量 V_n 依概率收敛到 0，而 c_nV_n 并不几乎必然趋于 0. [提示：令 A_n 为满足 $P(A_n)=1/n$ 的独立事件. 在 A_n 上，令 $V_n=1/c_n$，其他处为 0.]
7. 为何不用博雷尔–坎泰利引理证明强大数定律，并证明对 $\forall \varepsilon>0$，$\sum_n P(|S_n/n|>\varepsilon)<\infty$？因为后一个级数一般是发散的：令 $X_1, X_2, \cdots$ 为独立同分布的，具有密度 f，f 这样给出，当 $|x|\geqslant 1$ 时，$f(x):=|x|^{-3}$，当其他情况时，$f(x)=0$，所以对每一个博雷尔集 A，$P(X_j\in A)=\int_A f\mathrm{d}x$ 且 $EX_1=0$. 对 $n=1, 2, \cdots$，$j=1, \cdots, n$，令 A_{nj} 为事件 $\{X_j>n\}$，B_{nj} 为事件 $\left\{\sum_{1\leqslant i\leqslant n} X_i \geqslant X_j\right\}$. 令 $C_{nj}:=A_{nj}\cap B_{nj}$.
   (a) 证明：对每一个 $n\geqslant 2$ 和 j，$P(C_{nj})=n^{-2}/4$.
   (b) 证明：对每一个 n，$\{S_n/n>1\}\supset \bigcup_{1\leqslant j\leqslant n} C_{nj}$ 且 $\sum_n P(S_n>n)$ 发散. [提示：利用 Bonferroni 不等式和 $P(C_{ni}\cap C_{nj})\leqslant P(A_{ni}\cap A_{nj})$.]

266

*8.4 遍历定理

大数定律，或类似的事实，在关于变量没有独立性或正交性，且测度空间不需要是有限的条件下仍然可以证明. 设 $(X, \mathcal{A}, \mu)$ 为一个 σ-有限测度空间. 令 T 为一个从 X 映射到其自身的可测变换（= 函数），即对每一个 $B\in\mathcal{A}$，$T^{-1}(B)\in\mathcal{A}$. T 称为*保测的*（measure-preserving）当且仅当对所有的 $B\in\mathcal{A}$，$\mu(T^{-1}(B))=\mu(B)$. 集合 Y 称为 *T-不变的*（T-invariant）当且仅当 $T^{-1}(Y)=Y$. 因为 T^{-1} 保持所有的集合运算（比如并和补），所有 T-不变集的集族是一个 σ-代数. 因为任意两个 σ-代数的交集是一个 σ-代数，所有可测 T-不变集的集族 $\mathcal{A}_{\mathrm{inv}(T)}$ 是一个 σ-代数. 保测变换 T 称为*遍历的*（ergodic），当且仅当对每一个 $Y\in\mathcal{A}_{\mathrm{inv}(T)}$，或者 $\mu(Y)=0$，或者 $\mu(X\setminus Y)=0$.

例：

1. 对于 $\mathbb{R}$ 上的勒贝格测度，每一个变换 $x\mapsto x+y$ 是保测的（但并不是遍历的：见习题 1）.
2. 令 X 是平面上的单位圆，$X=\{(\cos\theta, \sin\theta): 0\leqslant\theta<2\pi\}$，其中测度 μ 由 $\mathrm{d}\mu(\theta)=\mathrm{d}\theta$ 给出，则 X 上的由一个角 α 给出的任意旋转 T 是保测的（对某个 α 遍历但对其他的则不遍历：见习题 4）
3. 对于 $\mathbb{R}^k$ 上的勒贝格测度 λ^k，平移、绕任意轴的旋转和任意平面中的反射是保测的. 这些变换并不是遍历的. 例如，对于 $k=2$，绕原点的旋转不是遍历的：环形 $a<r<b$ 对于极坐标 r 是不变量集合，其中这个集合干扰了遍历性.
4. 如果取一个概率空间 $(X, \mathcal{S}, P)$ 的副本的一个有限积，如同在 8.2 节中那样，则"移位"变换 $\{x_n\}_{n\geqslant1}\mapsto\{x_{n+1}\}_{n\geqslant1}$ 是保测的.（定理 8.4.5 中将证明它是遍历的.）
5. 保测变换不是一对一的一个例子：在有勒贝格测度的单位区间 $[0, 1]$ 上，当 $0\leqslant x<1/2$ 时，令 $f(x)=2x$；当 $1/2\leqslant x<1$ 时，令 $f(x)=2x-1$.

设 $f\in\mathcal{L}^1(X, \mathcal{A}, \mu)$，换言之，$f$ 是 X 上的一个可测实值函数且 $\int|f|\,\mathrm{d}\mu<\infty$. 对一个保测变换 T，令 $f_0:=f$ 且 $f_j:=f\circ T^j$，$j=1, 2, \cdots$（其中指数 j 表示合成，$T^j=T\circ T\circ\cdots\circ T$，有 j 项）. 令 $S_n:=f_0+f_1+\cdots+f_{n-1}$，则可以证明 S_n/n 是收敛的；如果 T 是遍历的，极限就是一个常数（正如在大数定理中那样）.

267

8.4.1 遍历定理 设 T 为一个 σ-有限测度空间 $(X, \mathcal{A}, \mu)$ 的任意保测变换，则对于任意

$f\in\mathcal{L}^1(X,\mathcal{A},\mu)$，存在一个函数 $\varphi\in\mathcal{L}^1(X,\mathcal{A}_{\mathrm{inv}(T)},\mu)$，使得对 μ-几乎所有的 x，$\lim_{n\to\infty}S_n(x)/n=\varphi(x)$，有 $\int|\varphi|\,\mathrm{d}\mu\leqslant\int|f|\,\mathrm{d}\mu$. 如果 T 是遍历的，则 φ(几乎处处)等于某个常数 c. 如果 $\mu(x)<\infty$，则 S_n/n 在 $\mathcal{L}^1$ 中收敛到 φ，即 $\lim_{n\to\infty}\int|\varphi-S_n/n|\,\mathrm{d}\mu=0$，所以 $\int\varphi\mathrm{d}\mu=\int f\mathrm{d}\mu$.

遍历定理的证明将依赖于另一个事实. 一个从 $\mathcal{L}^1(X,\mathcal{A},\mu)$ 映射到其自身的函数 U 称为线性的(linear)，当且仅当对任意的实数 c 和 $\mathcal{L}^1$ 中的函数 f 和 g，$U(cf+g)=cU(f)+U(g)$. 它称为正的(positive)，当且仅当 $f\geqslant0$(即对所有的 x，$f(x)\geqslant0$)，且 $f\in\mathcal{L}^1$，有 $Uf\geqslant0$. U 称为一个压缩(contraction)，当且仅当对所有的 $f\in\mathcal{L}^1$，$\int|Uf|\,\mathrm{d}\mu\leqslant\int|f|\,\mathrm{d}\mu$. 对任意一个从 X 映射到其自身的保测变换 T，$Uf:=f\circ T$ 定义了一个正线性压缩 U(由像测度定理4.1.11).

给定这样一个 U 和一个 $f\in\mathcal{L}^1$，令 $S_0(f):=0$，$S_n(f):=f+Uf+\cdots+U^{n-1}(f)$，$n\geqslant1$，令 $S_n^+(f):=\max_{0\leqslant j\leqslant n}S_j(f)$.

8.4.2 极大遍历引理 对 $\mathcal{L}^1(X,\mathcal{A},\mu)$ 上的任意正线性压缩 U，任意 $f\in\mathcal{L}^1(X,\mathcal{A},\mu)$，$n=0,1,\cdots$ 令 $A:=\{x:S_n^+(f)>0\}$，则 $\int_A f\mathrm{d}\mu\geqslant0$.

证明 对 $r=0,1,\cdots f+US_r(f)=S_{r+1}(f)$. 注意到 U 是正线性的，$g\geqslant h$ 意味着 $Ug\geqslant Uh$. 因此对 $j=1,\cdots,n$，$S_j(f)=f+US_{j-1}(f)\leqslant f+US_n^+(f)$.

如果 $x\in A$，那么 $S_n^+(f)(x)=\max_{1\leqslant j\leqslant n}S_j(f)(x)$. 结合对所有的 $x\in A$ 给出，$f\geqslant S_n^+(f)-US_n^+(f)$. 因为在 X 上 $S_n^+(f)\geqslant0$ 且在 A 之外 $S_n^+=0$，又因为 μ 是一个压缩，所以有

$$\begin{aligned}\int_A f\mathrm{d}\mu&\geqslant\int_A(S_n^+(f)-US_n^+(f))\,\mathrm{d}\mu=\int_X S_n^+(f)\,\mathrm{d}\mu-\int_A US_n^+(f)\,\mathrm{d}\mu\\&\geqslant\int_X S_n^+(f)\,\mathrm{d}\mu-\int_X US_n^+(f)\,\mathrm{d}\mu\geqslant0\end{aligned}$$

□

现在证明遍历定理(8.4.1)，对任意的实数 $a<b$，令

$$Y:=Y(a,b):=\{x:\liminf_{n\to\infty}S_n/n<a<b<\limsup_{n\to\infty}S_n/n\},$$

则 Y 是可测的. 如果在定义中用 $S_n\circ T$ 来代替 S_n 可得，$T^{-1}(Y)$ 取代了 Y. 现在 $S_n\circ T=S_{n+1}-f$，且当 $n\to\infty$ 时，$f/n\to0$，所以可用 S_{n+1} 代替 $S_n\circ T$. 但是在 Y 的原始定义中，S_n/n 能等价地由 $S_{n+1}/(n+1)$ 代替，然后由 S_{n+1}/n 代替. 所以 Y 是不变集.

268

为证明 $\mu(Y)=0$，假定 $b>0$，因为否则 $a<0$，并且我们可以考虑用 $-f$ 和 $-a$ 代替 f 和 b. 设 $C\subset Y$，$C\in\mathcal{A}$，且 $\mu(C)<\infty$. 令 $g:=f-b1_C$，$A(n):=\{x:S_n^+(g)(x)>0\}$，则由极大遍历引理(8.4.2)，对 $n=1,2,\cdots$

$$\int_{A(n)}f-b1_C\mathrm{d}\mu\geqslant0.$$

因为 $g\geqslant f-b$，对所有的 j 和 x，$S_j(g)\geqslant S_j(f)-jb$. 在 Y 上，$\sup_n(S_n(f)-nb)>0$. 因此 $Y\subset A:=\bigcup_n A(n)$. 因为当 $n\to\infty$ 时，$A(n)\uparrow A$，由控制收敛定理得，$\int_A f-b1_C\mathrm{d}\mu\geqslant0$，所以 $\int_A f\mathrm{d}\mu\geqslant b\mu(A\cap C)=b\mu(C)$. 因为 μ 是 σ-有限的，令 C 增加到 Y 并且得到 $\int f^+\mathrm{d}\mu\geqslant b\mu(Y)$. 由 $f\in\mathcal{L}^1$ 可得，$\mu(Y)<\infty$.

现在因为 Y 是不变的，我们可对 Y 随意限定且假设 $X=Y=C=A$. 那么有一有限测度空间，

$f-b\in\mathcal{L}^1$ 且 $\int_Y f-b\mathrm{d}\mu\geqslant 0$. 对 $a-f$, 由极大遍历引理和控制收敛定理可得, $\int_Y a-f\mathrm{d}\mu\geqslant 0$. 总之可得 $\int_Y a-b\mathrm{d}\mu\geqslant 0$. 因为 $a<b$, 所以 $\mu(Y)=0$.

取所有的有理数对 $a<b$, $S_n(f)/n$ 对于 μ 几乎处处(a. e.)收敛到某个在 $[-\infty,+\infty]$ 取值的函数 φ. T^n 对每一个 n 是保测的, 且

$$\int|f\circ T^n|\,\mathrm{d}\mu=\int|f|\circ T^n\mathrm{d}\mu=\int|f|\,\mathrm{d}(\mu\circ(T^n)^{-1})=\int|f|\,\mathrm{d}\mu,$$

这里中间的等式是由像测度定理(4.1.11)给出的. 进而得 $\int|S_n/n|\,\mathrm{d}\mu\leqslant\int|f|\,\mathrm{d}\mu$. 因此由法图引理(4.3.3), $\int|\varphi|\,\mathrm{d}\mu\leqslant\int|f|\,\mathrm{d}\mu<\infty$. 所以 φ 是几乎处处有限的, 特别地, 如果 φ 是有限的, $\varphi:=\limsup_{n\to\infty}S_n(f)/n$; 否则, 令 $\varphi=0$. 那么 φ 是可测的. 同样 φ 是 T-不变的, $\varphi=\varphi\circ T$, 正如在证明中 Y 是 T-不变的那样. 因此, 对于实数的任意博雷尔集 B, $\varphi^{-1}(B)=(\varphi\circ T)^{-1}(B)=T^{-1}(\varphi^{-1}(B))$, 所以 $\varphi^{-1}(B)\in\mathcal{A}_{\mathrm{inv}(T)}$ 且 φ 对于 $\mathcal{A}_{\mathrm{inv}(T)}$ 是可测的, 定理 8.4.1 中的第一个结论已证.

如果 T 是遍历的, 设 D 为满足 $\mu(\{x:\varphi(x)>y\})>0$ 的 y 的集合. 令 $c:=\sup D$, 则对于 $n=1$, 2, $\cdots\mu(\{x:\varphi(x)>c+1/n\})=0$, 所以 $\varphi\leqslant c$ a. e. 另一方面, 取 $y_n\in D$, 满足 $y_n<c$ 且 $y_n\uparrow c$. 对每一个 n, 因为 T 是遍历的, $\mu(\{x:\varphi(x)\leqslant y_n\})=0$, 所以 $\varphi\geqslant c$ a. e., 且 $\varphi=c$ a. e.. 因此 c 是有限的, 且如果 $\mu(X)=+\infty$, 则 $c=0$.

如果 $\mu(X)<\infty$ (T 不必是遍历的), 只需证明 $\int|\varphi-S_n/n|\,\mathrm{d}\mu\to 0$. 如果 f 是有界的, 如 $|f|\leqslant K$ a. e., 则对于所有的 n, $|S_n/n|\leqslant K$ a. e., 由控制收敛定理(定理 4.3.5)结论得证. 对于 $\mathcal{L}^1$ 中更一般的 f, 令 $f_K:=\max(-K,\min(f,K))$, 则每一个 f_K 是有界的, 且当 $K\to\infty$ 时, $\int|f_K-f|\,\mathrm{d}\mu\to 0$. 给定 $\varepsilon>0$, 令 $g=f_K$ 对足够大的 K 有 $\int|f-g|\,\mathrm{d}\mu<\varepsilon/3$. 那么 $S_n(g)/n$ 几乎处处收敛且收敛到 $\mathcal{L}^1$ 中的某个函数 j. 对于所有的 n, $\int|S_n(f)-S_n(g)|/n\mathrm{d}\mu\leqslant\varepsilon/3$. 由法图引理(4.3.3), $\int|\varphi-j|\,\mathrm{d}\mu\leqslant\varepsilon/3$. 对于足够大的 n, $\left|\int S_n(g)/n-j\right|\mathrm{d}\mu<\varepsilon/3$, 所以 $\limsup_{n\to\infty}\int|S_n(f)/n-\varphi|\,\mathrm{d}\mu\leqslant\varepsilon$. 令 $\varepsilon\downarrow 0$ 得后面的 lim sup(且因此极限)是 0. 因为对所有的 n, $\int S_n(f)/n\mathrm{d}\mu=\int f\mathrm{d}\mu$, 有 $\int\varphi\mathrm{d}\mu=\int f\mathrm{d}\mu$, 定理(8.4.1)得证.

□

下面的推论可以用同样的方式证明.

8.4.3 推论 如果在定理 8.4.1 中 $\mu(X)<\infty$ (其中 T 不必是遍历的)且对某个 $p(1<p<\infty)$, 有 $f\in\mathcal{L}^p$, 即 $\int|f|^p\mathrm{d}\mu<\infty$, 则 $S_n(f)/n$ 收敛到 $\mathcal{L}^p$ 中的 φ. 换言之, 当 $n\to\infty$ 时,

$$\int|S_n(f)/n-\varphi|^p\mathrm{d}\mu\to 0.$$

接下来, 关于积概率的事实是有用的. 设 I 为一标集, 或是所有非负整数的集合 $\mathbb{N}$, 或是所有整数的集合 $\mathbb{Z}$. 对每个 $i\in I$, 令 (Ω_i,S_i,P_i) 为一个概率空间. 令 $\Omega:=\prod_{i\in I}\Omega_i=\{\{x_i\}_{i\in I}$: 对每个 i, $x_i\in\Omega_i\}$, 则对 $i=\mathbb{N}$, $\{x_i\}_{i\in I}=\{x_i\}_{i\geqslant 0}$ 或对 $I=\mathbb{Z}$, 双边序列 $\{\cdots,x_{-2},x_{-1},x_0,x_1,x_2,\cdots\}$. 在 Ω

上有积 σ-代数$\mathcal{S}$，最小的 σ-代数使每个坐标 x_i 是可测的，且积概率由定理 8.2.2 给出 $Q:=Q^{(I)}:=\prod_{i\in I}P_i$.

对 $n=1,2,\cdots$，令$\mathcal{B}_n$为 Ω 子集的最小 σ-代数，使得对所有满足 $|j|\leqslant n$ 的 j，x_j 是可测的. 令$\mathcal{B}^{(n)}$为最小的 σ-代数，使得对所有的 $j>n$，x_j 是可测的. 令$\mathcal{B}^{(\infty)}:=\bigcap_{n=1}^{\infty}\mathcal{B}^{(n)}$，则$\mathcal{B}^{(\infty)}$称为尾事件(tail event)$\sigma$-代数. 例如，如果对所有的 i，$\Omega_i=\mathbb{R}$，有博雷尔 σ-代数和 $S_n:=x_1+\cdots+x_n$，则事件$\{\limsup_{n\to\infty}S_n/n>0\}$在$\mathcal{B}^{(\infty)}$中.

8.4.4 科尔莫戈罗夫 0－1 律 对任意的积概率 Q 和 $A\in\mathcal{B}^{(\infty)}$，$Q(A)=0$ 或 1.

证明 令$\mathcal{C}$为$\mathcal{S}$中所有集合 D 构成的集族，使得对每个 $\delta>0$，存在某个 n 和 $B\in\mathcal{B}_n$，满足 $Q(B\Delta D)<\delta$，其中 $B\Delta D$ 是对称差$(B\setminus D)\cup(D\setminus B)$. 那么对所有的 n，$\mathcal{C}\supset\mathcal{B}_n$. 如果 $D\in\mathcal{C}$，则对于其补集 D^c 和任意的 B，有 $B^c\Delta D^c=B\Delta D$，所以 $D^c\in\mathcal{C}$. 对于任意 $D_j\in\mathcal{C}(j=1,2,\cdots)$，令 $D:=\bigcup_j D_j$. 给定 $\delta>0$，存在一个 m，使得 $Q\left(D\setminus\bigcup_{j=1}^{m}D_j\right)<\delta/2$. 对每个 $j=1,\cdots,m$，存在一个 $n(j)$和 $B_j\in\mathcal{B}_{n(j)}$，使得 $Q(D_j\Delta B_j)<\delta/(2m)$. 令 n: $\max_{j\leqslant m}n(j)$ 且 $B:=\bigcup_{j\leqslant m}B_j$. 那么 $B\in\mathcal{B}_n$且 $Q(D\Delta B)<\delta$，所以 $D\in\mathcal{C}$. 故$\mathcal{C}$是一个 σ-代数，$\mathcal{C}=\mathcal{S}$.

270

给定 $A\in\mathcal{B}^{(\infty)}$，取 $B_n\in\mathcal{B}_n$且当 $n\to\infty$ 时，$Q(B_n\Delta A)\to0$. 因为 Q 是一个积概率且 $A\in\mathcal{B}^{(n)}$，对所有的 n，$Q(B_n\cap A)=Q(B_n)Q(A)$. 令 $n\to\infty$ 得 $Q(A)=Q(A)^2$，故 $Q(A)=0$ 或 1. □

现在考虑特殊的情况，其中所有的$(\Omega_i,\mathcal{S}_i,P_i)$是一个概率空间$(X,\mathcal{A},P)$的副本. 那么令 $X^I:=\Omega$ 和 $P^I:=Q$，使得坐标 x_i 是 i. i. d. (P). 移位变换(shift transformation)T 定义为 $T(\{x_i\}_{i\in I}):=\{x_{i+1}\}_{i\in I}$，则 T 是一个对于测度 P^I 由 X^I 映上到其自身的保测变换，其中 $I=\mathbb{N}$ 为单侧移位(unilateral shift)和 $I=\mathbb{Z}$ 为双向移位(bilateral shift).

8.4.5 定理 对 $I=\mathcal{N}$或$\mathbb{Z}$ 和任意的概率空间$(X,\mathcal{A},P)$，移位 T 对于 P^I 在 X^I 上总是遍历的.

证明 如果 $I=\mathbb{N}$，则对任意的 $Y\in\mathcal{B}$，$T^{-1}(Y)\in\mathcal{B}^{(0)}$，$T^{-1}(T^{-1}(Y))\in\mathcal{B}^{(1)}$，等等，所以如果 Y 是一不变集，则 $Y\in\mathcal{B}^{(\infty)}$. 那么 0-1 律(8.4.4)意味着 T 是遍历的.

如果 $I=\mathbb{Z}$，则给定 $A\in\mathcal{B}$，像在 8.4.4 的证明中那样，取 $B_n\in\mathcal{B}_n$，当 $n\to\infty$ 时，$P^I(B_n\Delta A)\to0$. 如果 A 是不变的，$A=T^{-2n-1}A$，所以

$$P^I(T^{-2n-1}(B_n\Delta A))=P^I(T^{-2n-1}B_n\Delta A)\to0.$$

另一方面，$T^{-2n-1}B_n\in\mathcal{B}^{(n)}$，所以 $T^{-2n-1}B_n$ 和 B_n 是独立的. 令 $n\to\infty$，则同样给出 $P^I(A)^2=P^I(A)$，$P^I(A)=0$ 或 1. □

对于有有限均值的独立同分布变量 X^n，遍历定理给出了强大数定律(8.3.5)的另一个证明如下. 设 I 为正整数的集合. 考虑从 Ω 映射到$\mathbb{R}^I$ 定义为 $Y(\omega):=\{X_n(\omega)\}_{n\geqslant1}$ 的函数 Y，则像测度 $P\circ Y^{-1}$等于 Q^I，其中 $Q:=\mathcal{L}(X_1)$，由定理 8.2.2 中的独立性和唯一性，可以假设 $\Omega=\mathbb{R}^I$，X_n 为坐标，且 $P=Q^I$. 由定理 8.4.5 知，T 是遍历的，由遍历定理(8.4.1)得，独立同分布强大数定律(8.3.5)成立.

积 X^I 中的集合 A 称为对称的，如果对任意变换 T，$T(A)=A$，这里 T 是从 I 映上到其自身的变换，对于某个 1－1 函数 π，满足 $T(\{x_n\}_{n\in I})=\{x_{\pi(n)}\}_{n\in I}$，其中对几乎有限多个$j$，$\pi(j)=j$. 那么 π 称为一个有限置换.

271

8.4.6 定理(Hewitt-Savage 0－1 律)　设 I 为任意可数无限指标集，A 为积空间 X^I 中的任意可测对称集，且 P 为 X 上的任意概率测度，则 $P^I(A)=0$ 或 1.

证明　假设 $I=\mathbb{Z}$. 对每个 $n=1, 2, \cdots$ 和 $j\in\mathbb{Z}$，如果 $|j|>n$，令 $\pi_n(j)=j$，如果 $-n\leqslant j<n$，令 $\pi_n(j)=j+1$，且 $\pi_n(n):=-n$. 那么 π_n 是一个有限置换. 令 $\mathcal{F}_n$ 为最小的 σ-代数，其中 X_i 对于 $|i|\leqslant n$ 是可测的. 给定 $\varepsilon>0$，如同在 8.4.4 的证明中那样，存在 n 和 $B_n\in\mathcal{F}_n$，满足 $P^I(A\Delta B_n)<\varepsilon/2$. 令 $\zeta_n(x):=\{x_{\pi_n(j)}\}_{j\in\mathbb{Z}}$. 对所有的 j，令 $\pi_\infty(j):=j+1$，故 ζ_∞ 是移位. 每个 ζ_n 对 P^I 和 A 是不变的，所以 $\varepsilon/2>P^I(A\Delta\zeta_{n+2}(B_n))=P^I(A\Delta\zeta_\infty(B_n))=P^I(\zeta_\infty^{-1}A\Delta B_n)$，且 $P^I(\zeta_\infty^{-1}(A)\Delta A)<\varepsilon$，因此 $P^I(\zeta_\infty^{-1}A\Delta A)=0$. 由下面的习题 7 知，对于一个不变集 B，$P^I(A\Delta B)=0$，且由定理 8.4.5 得，$P(B)=0$ 或 1，并且对于 A 结论同样成立. □

习题

1. 在具有勒贝格测度的$\mathbb{R}$上，证明：保测变换 $x\mapsto x+y$ 不是遍历的.
2. 证明推论 8.4.3.
3. 在 $X=\mathbb{R}^{\mathbb{N}}$ 或$\mathbb{R}^{\mathbb{Z}}$上，设 A 为所有序列$\{x_n\}$的集合，当 $n\to+\infty$ 时，满足$(x_1+\cdots+x_n)/n\to 1$. 证明：A
   (a)对于$\mathbb{R}^{\mathbb{N}}$ 或$\mathbb{R}^{\mathbb{Z}}$ 中的单侧移位$\{x_j\}\mapsto\{x_{j+1}\}$是否是不变的.
   (b)在$\mathbb{R}^{\mathbb{N}}$ 或$\mathbb{R}^{\mathbb{Z}}$ 中是否是对称的(如同在 Hewitt-Savage 0－1 律中那样).
4. 设 S^1 为$\mathbb{R}^2$ 中的单位圆，$S^1:=\{(x, y): x^2+y^2=1\}$，且对于通常的极坐标 θ，弧长测度为 μ: $\mathrm{d}\mu(\theta)=\mathrm{d}\theta$. 令 T 为由角 α 给出的 S^1 的旋转. 证明：T 是遍历的当且仅当 α/π 是无理数. [提示：如果 $\varepsilon>0$ 且 $\mu(A)>0$，由命题 3.4.2 知，有一个满足 $\mu(A\cap J)>(1-\varepsilon)\mu(J)$ 的区间 J: $a<\theta<b$. 如果 α/π 是无理数，证明对任意 $z\in S^1$，$\{T^n z\}_{n\geqslant 0}$在 S^1 中稠密.]
5. (a)在具有计数测度的$\mathbb{Z}$上，证明：移位变换 $n\mapsto n+1$ 是遍历的.
   (b)令 π 为从$\mathbb{N}$映上到其自身的 1－1 函数，它是$\mathbb{N}$上计数测度的一个遍历变换. 令$(X, \mathcal{B}, P)$为一个概率空间. 从$\mathbb{X}^{\mathbb{N}}$映上到其自身的 T_π定义为 $T_\pi(\{x_i\}_{i\in\mathbb{N}}):=\{x_{\pi(i)}\}_{i\in\mathbb{N}}$. 证明：$T_\pi$是遍历的.
6. 在单位区间[0, 1)上，取勒贝格测度. 由它们的十进制展开 $x=0.x_1x_2x_3\cdots$来表示数字，其中每个 x_j 是从 0～9 的一个整数，且对有两个展开的这些数，以一个无穷的串 0 或 9 结尾，选择以 0 结尾的.
   (a)证明：数字的移位变换$\{x_n\}\mapsto\{X_{n+1}\}$是保测的.
   (b)证明：这个变换是遍历的.

272

7. 给定一个 σ-有限测度空间$(X, \mathcal{S}, \mu)$和一个从 X 映射到其自身的保测变换 T，集合 $A\in\mathcal{S}$称为殆不变的(almost invariant)当且仅当$\mu(A\Delta T^{-1}(A))=0$. 证明：对于任意殆不变集 A，有一个不变可测集 B，满足 $\mu(A\Delta B)=0$. [提示：对于 $n\geqslant 1$，令 $T^{-n-1}(A):=T^{-1}(T^{-n}(A))$且 $B:=\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}T^{-n}(A)$.]
8. 对于一个 σ-有限测度空间$(X, \mathcal{S}, \mu)$，如果 T 是 X 上的一个遍历保测变换，证明：T 对于 μ 的完备化有同样的性质，其中 σ-代数$\mathcal{S}$扩张到包含所有 μ 零测集的子集.
9. 对 $j=1, 2$，设$(X_j, \mathcal{S}_j, \mu_j)$为一个 σ-有限测度空间，令 T_j 为 X_j 的保测变换. 令 $T(x_1, x_2):=(T_1(x_1), T_2(x_2))$.
   (a)证明：T 是$(X_1\times X_2, \mathcal{S}_1\oplus\mathcal{S}_2, \mu_1\times\mu_2)$的一个保测变换.
   (b)证明：即使 T_1 和 T_2 都是遍历的，T 也可能不是遍历的. [提示：令每个 X_j 为圆且令每个 T_j 通过相同的角 α 作旋转(如同在习题 4 中那样)且 α/π 是无理数.]

注释

8.1 节　对于一个均匀的骰子，每一面的概率恰为 1/6，我们可得 $P(\{5, 6\})=1/3$. 在实际使

用骰子时，每一面上的数字用挖空的牌点来标记. 6 那一面是最轻的，相反 1 那一面是最不轻的. 类似地，5 那一面比它的背面 2 那一面要轻. 一些实际的试验给出 $P(\{5, 6\}) = 0.3377 \pm 0.0008$ （“Weldon 的骰子数据”；见 Feller，1968，p. 148-149）.

集合函数 P 称为*有限可加的*，当且仅当对于任意有限的 n 和 P 的定义域中的不交集 $A_1, \cdots, A_n$，

$$P(A_1 \cup \cdots \cup A_n) = \sum_{j=1}^{n} P(A_j).$$

3.1 节末尾的一些习题表明病态的发生仅是有限可加而不是可数可加. 概率定义为一般空间中 σ-代数上的全质量为 1 的（可数可加、非负）测度，被概率论方面的大多数研究人员所采用. Kolmogorov(1933) 的著作第一次使这个定义为人所周知（例如，见 Bingham(2000)）. Barone 和 Novikoff(1978) 这篇论文（他们计划的部分 Ⅱ 显然还没有出现）首先涉及了 Borel(1909)，其次涉及了在 Hausdorff(1914) 中的一些概率的例子.

Kolmogorov(1929b) 给出了在一般空间上概率的公理化，从一些集合（这些集合并不需要是代数）构成的集族上的一个有限可加函数开始，比如整数集合的密度（3.1 节习题 9）. 在公理 Ⅰ 中，概率要求是“ >0 ”（可能意指“ $\geqslant 0$ ”）. 一个“正态的”概率定义在一个代数上是可数可加的，且因此可唯一地扩张到一个 σ-代数上，如在定理 3.1.10 中那样. 但是 Kolmogorov(1929b) 仍然没有限制概率是“正态的”，这篇论文几乎没有受到关注，也没有列到当时的数学评论杂志《Jahrbuch über die Fortschritte der Mathematik》或《Revue Semestrielle des Publications Mathématiques》中. 这些评论杂志，甚至在更吸引数学家注意的杂志中，几乎很少包含俄语的论文. Kolmogorov(1929b) 被收入在 Kolmogorov(1986) 而且之后由 Shiryayev(1989，p. 884) 提及. [273]

因此，Ulam(1932) 可能是最早向国际听众给出了“Kolmogorov(1933)”概率的定义，显然与 Kolmogorov 无关. 此外，Ulam 还要求单元素集是可测的. Ulam 的注释是由 Lomnicki 和 Ulam(1934) 联合发表的文章的声明，其中引用了 Kolmogorov(1933). 在先前的 20 年里，概率测度是定义在特殊的空间（比如，欧几里得空间）上并且需要进一步的限制条件. 往往取具有勒贝格测度的 $[0, 1]$ 为基本的概率空间.

Kolmogorov(1933) 还包括很重要的一个条件期望的定义和一个随机过程中的存在性定理（见 10.1 节和 12.1 节的注释）. Kolmogorov 的其他著作在这章或后面的章节中都会提到. 正如 Gnedenko and Smirnov(1963) 写到，“在当代的概率论中，A. N. Kolmogorov 当然是公认的领袖.” 1980 年，Kolmogorov（和 Henri Cartan）获得沃尔夫数学奖，第三年授予了这个奖（注意，1981 年）. 颂扬他的其他文章是 Gnedenko(1973) 和 The Times' obituary(1987).

Dubins 和 Savage(1965) 认为概率仅仅是有限可加的. 有限可加“概率”的其他发展引入了不同的方向，甚至在文献中往往不称为概率.

一个主要的例子是“不变平均”. 令 G 为一个阿贝尔群：即从 $G \times G$ 映射到 G 定义了一个函数 $+$，对 G 中的任意 x, y, z，$x + y = y + x$ 且 $x + (y + z) = (x + y) + z$，且存在一个 $0 \in G$，使得对每个 $x \in G$，$x + 0 = x$ 且存在一个 $y \in G$，满足 $x + y = 0$. 那么不变平均（invariant mean）就是一个定义在 G 的所有子集上的函数 d，非负且有限可加，满足 $d(G) = 1$，对于所有的 $A \subset G$ 和 $m \in G$，$d(A + m) = d(A)$，其中 $A + m := \{a + m : a \in A\}$. 不变平均也可以定义在半群上，比如所有正整数的集合 $\mathbb{Z}^+$. 在 $\mathbb{N}$ 或 $\mathbb{Z}^+$ 上，不变平均是“密度”的扩张，见 3.1 节的习题 9. 关于不变平均理论，可以

参考 Banach(1923)、von Neumann(1929)、Day(1942)和 Greenleaf(1969).

投针到直纹平面的问题与 Buffon(1733, p. 43—45; 1778, p. 147—153)有关，可以参考 Jordan(1972, p. 23). 顺便说一下，Jordan 在法国用名字 Charles Jordan 发表了许多论文；在德国用名字 Karl Jordan 发表了一些论文；在匈牙利用名字 Jordan Károly 发表了多篇论文.

8.2 节 P. J. Daniell(1919)首先证明了在单位区间[0, 1]上每个因式是勒贝格测度 λ 的情况下，在无穷积上可数可加积概率的存在性. Daniell 的证明运用他的方法到积分(4.5 节)、紧性和迪尼定理(2.4.10)，以仅依赖于有限多个坐标的连续函数开始，正如在 Daniell 积分理论中的$\mathcal{L}$类.

274

对于一个可测函数 f，在应用中用到的几乎所有概率测度都能表示为勒贝格测度的像 $\lambda \circ f^{-1}$. 根据 Daniell 定理人们很容易得到这些像的积. Kolmogorov(1933)构造了积空间(比如实轴或者局部紧空间)上的概率(不仅仅是积概率)，见 12.1 节.

显然 Lomnicki 和 Ulam(1934)首先证明了一个没有任何拓扑的任意概率空间积的存在性(定理 8.2.2). Ulam(1932)已经宣布这是他和 Lomnicki. J 共同证明的一个结果. 显然 von Neumann(1935)独立地证明了这个定理，这可从(普林斯顿)高等研究中心的讲稿中看出，这份讲稿直到 1950 年才发表. 与此同时，B. Jessen(1939)在其他人(包括 E. S. Andersen，他给出了错误的证明)之后用丹麦语发表了这个定理. Kakutani(1943)也证明了这个结果. 最后 Andersen 和 Jessen(1946, p. 20)用英语发表了 Jessen 的证明，并且这个定理通常称为 Andersen-Jessen 定理. 对积上测度的扩张其中在 X_n 上的测度依赖于 $x_1, \cdots x_n$)通常归功于 Ionescu Tulcea(*1949—1950*). Andersen 和 Jessen(1948, p. 5)谈到 Doob 也意识到了这样一个结果并且 Doob 和 Jessen 也计划着写一篇关于这方面的论文，但是我并没有在《Mathematical Reviews》1940—1959 的索引中找到由 Doob 和 Jessen 联合发表的文章. 我感激 Lucien Le Cam 提供的一些参考文献.

8.3 节 Révész(1968)写一本关于大数定律的书. Gnedenko 和 Kolmogorov(1949, 1968)是有关极限定理的经典著作，其中还包括大数定律. 下面是一些众所周知的结果. 如果 $X_1, X_2, \cdots$是独立同分布的，则 Kolmogorov(1929a)的一个定理讲到，存在常数 a_n，满足 $S_n/n - a_n$ 依概率收敛到 0，当且仅当 $M \to +\infty$ 时，$MP(|X_1| > M) \to 0$(Révész, 1968, p. 51). 令 φ 为 X_1 的特征函数，$\varphi(t) := E\exp(\mathrm{i}X_1 t)$，则存在一个常数 c，使得 S_n/n 依概率收敛到 c，当且仅当 φ 在 0 处有导数，且 $\varphi'(0) = \mathrm{i}c$，Ehrenfeucht 和 Fisz(1960)的定理在 Révész(1968, p. 52)中也有介绍. 当弱大数定律成立时强大数定律可以不成立. 例如，对任意博雷尔集 $B \subset \mathbb{R}$，令

$$P(X_1 \in B) = \beta \int_B 1_{\{|x| \geq 2\}} x^{-2} (\log |x|)^{-\gamma} \mathrm{d}x,$$

其中选择常数 β 使得在整条线上的积分是 1. 如果 $X_1, X_2, \cdots$是独立同分布的，则强大数定律成立当且仅当 $\gamma > 1$. 弱大数定律，并非强大数定律，对于 $0 < \gamma \leq 1$ 成立，有 $c = 0$. 对于 $\gamma \leq 0$，弱大数定律不成立，即使有变量 a_n 也不成立，因为不满足科尔莫戈罗夫条件. 如果去掉对数因式而且对于某个 $p > 1$，用 $|x|^{-p}$代替 x^{-2}(对得到一个有限测度，然后得到一个概率测度，这是必需的)，则强大数定律对于 $p > 2$ 成立而且弱大数定律不恰对 $p \leq 2$ 成立. 所以弱大数定律，如果没有强大数定律，其应用范围很小.

Bienaymé(1853, p. 321)显然最早地给出了有些不完美的不等式形式(8.3.1)并且得出了弱大数定律. Chebyshev(1867, p. 183)证明了更加精确的结论. 他被誉为是第一个严格证明了概率中一般极限定理的人，他在数学和技术上还有一些其他成就(Youschkevitch, 1971). Bienaymé(1853, p. 315)还注意到对于独立变量，和的方差等于方差的和. 关于 Bienaymé 参见 Heyde 和 Seneta

275

(1977). 与 Bienayme-Chebyshev 有关的不等式已经被 Godwin(1955, 1964)和 Savage(1961)勘定，他们说他的参考书目是"尽可能完善的".

Jakob Bernoulli(1713)最先证明了弱大数定律，对于独立同分布的 X_j，每个仅取两个值——二项式或"伯努利"情况. Sheynin(1968)研究了弱大数定律早期历史.

Borel(1909, p. 252)在事件独立的情况下给出了博雷尔-坎泰利引理(8.3.4)，并给出了一个不充分的证明. Cantelli(1917a, 1917b)发现在不独立时有一半是成立的，如同在特殊的情况下的 Hausdorff(1914, p. 421). Barone 和 Novikoff(1978)指出博雷尔在概率论基础方面的工作具有里程碑的作用，但也对他的证明进行了批评. 他们注意到 Borel(1903, Thme. XI bis)近乎(在 ε 之内，几何概率的情况下)给出了"坎泰利"引理的一半. Cohn(1972)给出了博雷尔-坎泰利引理的一些扩张. Erdös 和 Rényi(1959)证明了博雷尔-坎泰利引理对两两独立的事件成立. 证明见 Chung(1974, 定理 4.2.5, p. 76). 关于博雷尔-坎泰利引理的历史见 Móri 和 Székely(1983). 坎泰利生于 1875 年卒于 1966 年. Benzi(1988)回顾了他的一些著作，尤其是概率论基础方面的.

定理 8.3.2 中的条件 $EX_j^2=1$ 可以由 $\sum_{j=1}^{n} EX_j^2/n^2 \to 0$ $(n\to\infty)$代替，如 A. A. Markov((1899)所注释的. 关于马尔可夫，以他的名字命名的马尔可夫链和马尔可夫过程，见 Youschkevitch(1974).

最早的强大数定律，也是在伯努利情况下，由 Borel(1909)讲述的，且同样没有一个正确的证明. 对于独立变量满足 $EX_n^4<\infty$ 的 X_n，且在更进一步的限制下，包括独立同分布变量，根据 Seneta(1992)的描述，Cantelli(1917a)证明了强大数定律，其中 Seneta(1992)讲述了更多的关于大数定律的发展历史.

Kolmogorov(1930)证明了如果 X_n 是独立随机变量且满足 $EX_n=0$ 和 $\sum_{n\geq 1} EX_n^2/n^2 < \infty$，则大数定律对于它们成立. Kolmogorov(1933)给出了定理 8.3.5，对于独立同分布随机变量，强大数定律成立当且仅当 $E|X_1|<\infty$. 1930 和 1933 中的定理之间的关系参见第 9 章的注释. 强大数定律 8.3.5 也可以利用科尔莫戈罗夫的另一个定理，由 G. D. Birkhoff(1932)和 Khinchin(1933)的遍历定理证得，如同在 8.4 节中证明的.

上面给出的强大数定律的简短证明应归功于 Etemadi(1981). 他叙述并证明了独立分布变量只需要是两两独立的结果. 在应用中，似乎很少遇到两两独立的而非独立的变量序列.

8.4 节 L. Boltzmann(1887, p. 208)是与统计力学相关联的. 对于在密闭容器中的 n 分子的气体，所有分子的可能位置和要素的集合是一个维数 $d=6n$ 的"相位空间". 如果旋转，等等，考虑单个分子，维数仍然很大. 令 $z(t)\in\mathbb{R}^d$ 为气体在时刻 t 的状态，则 $z(t)$ 在紧曲面 S 上对于所有的 t 保持状态，其中能量是常数(或许，一些其他的变量称为"运动的积分"). 根据经典力学，对于每个 t，有一个变换 T_t，对所有的 s 和所有可能的轨道 $z(\cdot)$，使得 $z(s)$ 映射到 $z(s+t)$. 对所有的保测变换 T_t，在 S 上存在一个概率测度 P. Boltzmann 的最初的"遍历假设"是 $z(t)$ 能够跑遍 S 中的所有点，除了在特殊情况下. 因此，对于 S 上的任意连续函数 f，一个长时间的平均将等于一个期望， 276

$$\lim_{T\to\infty} T^{-1}\int_0^T f(z(t))\,\mathrm{d}t = \int_S f(z)\,\mathrm{d}P(z).$$

但是 M. Plancherel(1912, 1913)和 A. Rosenthal(1913)独立地证明了一条光滑曲线(比如 $z(\cdot)$)并不能填满一个维数大于 1 的流形(尽管一条连续的佩亚诺曲线能够做到：见 2.4 节习题 9). 为了看到这一点，一种方式是把 z 限制到一个有限时间区间上，值域在 S 中无处稠密. 因为 S 是完备的，由

范畴定理(2.5.2)知，z的值域不可能取到S中的所有点．一般来说，关于贝尔(R. Baire)的工作，包括众所周知的由他的名字命名的范畴定理的应用，见 Dugac(1976)．另一种方法是证明z的值域在S上测度为0.

接着就要问是否z的值域在S上是稠密的，而且时间和空间的均值仍然是相等的．正如 Boltzmann 自己指出的，对于与S中的所有点相对接近的轨道可以取无度的长时间．Brush(1976，Book2，p. 363—385)回顾了遍历假设的历史．Brush(1976，Book1，p. 80)评价说遍历定理"现在看起来数学家比物理学家更感兴趣."见 D. ter Haar(1954，p. 123—125 和 331—385，特别是 356 ff).

遍历定理(8.4.1)是"空间"和长时间平均相等的离散时间版本．Birkhoff(1932)首先证明了它对于示性函数与一类特殊的测度空间和变换成立．Khinchin(1933)指出证明可以直接推广到任意有限测度空间上的$\mathcal{L}^1$函数．E. Hopf(1937)对于保测变换，证明了极大遍历引理(8.4.2)，通常称为极大遍历定理．Yosida 和 Kakutani(1939)证明了怎样用极大遍历引理证明遍历定理．Garsia(1965)给出了极大遍历引理的简短证明.

回忆$\mathcal{L}^\infty(X, \mathcal{A}, \mu)$，$X$上可测实函数$f$的空间且$\|f\|_\infty < \infty$，其中，

$$\|f\|_\infty := \inf\{M:\mu\{x: |f(x)| > M\} = 0\}.$$

$\mathcal{L}^1$中的压缩U称为是强的(strong)，当且仅当它也是$\mathcal{L}^\infty$半范数中的一个压缩，所以对所有的$f\in\mathcal{L}^1\cap\mathcal{L}^\infty$，$\|Uf\|_\infty \le \|f\|_\infty$．Dunford 和 Schwartz(1956)证明了如果U是一个强压缩，则当$n\to\infty$时，$n^{-1}\sum_{1\le i\le n} U^i f$几乎处处收敛．Chacon(1961)给出了一个简短的证明且 Jacobs(1963，pp. 371—376)给出另一个说明.

如果U是$\mathcal{L}^1$的任意正压缩且g是$\mathcal{L}^1$中的一个非负函数，则商$\sum_{0\le i\le n} U^i f \Big/ \sum_{0\le i\le n} U^i g$在一个集合上几乎处处收敛，其中对于某个$n$，分母是正的(且因此几乎处处，如果处处$g>0$)．Hopf(1937)证明了这个对于$Uf=f\circ T$成立，其中$T$是一个保测变换．Chacon 和 Ornstein(1960)给出了一般证明．他们的证明相当长(见 Jacobs，1963，p. 381—400)．Brunel(1963)和 Garsia(1967)缩短了这个证明．注意到如果$\mu(X)<\infty$，则可以令$g=1$，所以像前面的定理中那样，分母变成了n.

Halmos(1953)、Jacobs(1963)和 Billingsley(1965)给出了遍历定理的一般说明．后来，在保持测度不变进行的同构分类上取得了实质性的进展(见 Ornstein et al.，1982).

Hewitt 和 Savage(1955，定理 11.3)证明了他们的 0－1 律(8.4.6)．他们给出了三个证明，涉及 Halmos 已经告诉他们前面给出的简短证明.

277

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到.

Andersen, Eric Sparre, and Borge Jessen (1946). Some limit theorems on integrals in an abstract set. *Danske Vid. Selsk. Mat.-Fys. Medd.* 22, no. 14. 29 pp.

——— and ——— (1948). On the introduction of measures in infinite product sets. *Danske Vid. Selsk. Mat.-Fys. Medd.* 25, no. 4. 8 pp.

Banach, Stefan (1923). Sur le problème de la mesure. *Fundamenta Mathematicae 4*: 7–33.

Barone, Jack, and Albert Novikoff (1978). A history of the axiomatic formulation of probability from Borel to Kolmogorov: Part I. *Arch. Hist. Exact Sci. 18*: 123–190.

Benzi, Margherita (1988). A "neoclassical probability theorist:" Francesco Paolo Cantelli (in Italian). *Historia Math. 15*: 53–72.

Bernoulli, Jakob (1713, posth.). *Ars Conjectandi*. Thurnisiorum, Basel. Repr. in *Die Werke von Jakob Bernoulli 3* (1975), pp. 107–286. Birkhäuser, Basel.

Bienaymé, Irenée-Jules (1853). Considérations a l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C.R. Acad. Sci. Paris 37*: 309–324. Repr. in *J. math. pures appl.* (Ser. 2) *12* (1867): 158–276.

Billingsley, Patrick (1965). *Ergodic Theory and Information*. Wiley, New York.

Bingham, N. H. (2000). Studies in the history of probability and statistics XLVI. Measure into probability: From Lebesgue to Kolmogorov. *Biometrika 87*: 145–156.

Birkhoff, George D. (1932). Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA 17*: 656–660.

Boltzmann, Ludwig (1887). Ueber die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *J. für die reine und angew. Math. 100*: 201–212.

Borel, Émile (1903). Contribution à l'analyse arithmétique du continu. *J. math. pures appl.* (Ser. 2) *9*: 329–375.

——— (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti Circolo Mat. Palermo 27*: 247–271.

Brunel, Antoine (1963). Sur un lemme ergodique voisin du lemme de E. Hopf, et sur une de ses applications. *C. R. Acad. Sci. Paris 256*: 5481–5484.

Brush, Stephen G. (1976). *The Kind of Motion We Call Heat*. 2 vols. North-Holland, Amsterdam.

*Buffon, Georges L. L. (1733). *Histoire de l' Académie*. Paris.

*——— (1778). X Essai d'Arithmétique Morale. In *Histoire Naturelle*, pp. 67–216. Paris.

*Cantelli, Francesco Paolo (1917a). Sulla probabilità come limite della frequenza. *Accad. Lincei, Roma, Cl. Sci. Fis., Mat., Nat., Rendiconti* (Ser. 5) *26*: 39–45.

*——— (1917b). Su due applicazione di un teorema di G. Boole alla statistica matematica. Ibid., pp. 295–302.

Chacon, Rafael V. (1961). On the ergodic theorem without assumption of positivity. *Bull. Amer. Math. Soc. 67*: 186–190.

——— and Donald S. Ornstein (1960). A general ergodic theorem. *Illinois J. Math. 4*: 153–160.

Chebyshev, Pafnuti Lvovich (1867). Des valeurs moyennes. *J. math. pures appl. 12* (1867): 177–184. Transl. from *Mat. Sbornik 2* (1867): 1–9. Repr. in *Oeuvres de P. L. Tchebychef 1*, pp. 687–694. Acad. Sci. St. Petersburg (1899–1907).

Chung, Kai Lai (1974). *A Course in Probability Theory*. 2d ed. Academic Press, New York.

Cohn, Harry (1972). On the Borel-Cantelli Lemma. *Israel J. Math. 12*: 11–16.

Daniell, Percy J. (1919). Integrals in an infinite number of dimensions. *Ann. Math. 20*: 281–288.

Day, Mahlon M. (1942). Ergodic theorems for Abelian semi-groups. *Trans. Amer. Math. Soc. 51*: 399–412.

Dubins, Lester E., and Leonard Jimmie Savage (1965). *How to Gamble If You Must; Inequalities for Stochastic Processes*. McGraw-Hill, New York.

Dugac, P. (1976). Notes et documents sur la vie et l'oeuvre de René Baire. *Arch. Hist. Exact Sci. 15*: 297–383.

Dunford, Nelson, and Jacob T. Schwartz (1956). Convergence almost everywhere of operator averages. *J. Rat. Mech. Anal.* (Indiana Univ.) *5*: 129–178.

Ehrenfeucht, Andrzej, and Marek Fisz (1960). A necessary and sufficient condition for the validity of the weak law of large numbers. *Bull. Acad. Polon. Sci. Ser. Math. Astron. Phys. 8*: 583–585.

*Erdös, Paul, and Alfréd Rényi (1959). On Cantor's series with divergent $\sum 1/q_n$. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* 2: 93–109.

Etemadi, Nasrollah (1981). An elementary proof of the strong law of large numbers. *Z. Wahrsch. verw. Geb. 55*: 119–122.

Feller, William (1968). *An Introduction to Probability Theory and Its Applications.* Vol. 1, 3d ed. Wiley, New York.

Garsia, Adriano M. (1965). A simple proof of E. Hopf's maximal ergodic theorem. *J. Math. and Mech.* (Indiana Univ.) *14*: 381–382.

——— (1967). More about the maximal ergodic lemma of Brunel. *Proc. Nat. Acad. Sci. USA 57*: 21–24.

Gnedenko, Boris V. (1973). Andrei Nikolaevich Kolmogorov (On his 70th birthday). *Russian Math. Surveys 28*, no. *5*: 5–17. Transl. from *Uspekhi Mat. Nauk 28*, no. 5: 5–15.

——— and Andrei Nikolaevich Kolmogorov (1949). *Limit Distributions for Sums of Independent Random Variables.* Translated, annotated, and revised by Kai Lai Chung, with appendices by Joseph L. Doob and Pao Lo Hsu. Addison-Wesley, Reading, Mass. 1st ed. 1954, 2d ed. 1968.

——— and N. V. Smirnov (1963). On the work of A. N. Kolmogorov in the theory of probability. *Theory Probability Appls. 8*: 157–164.

Godwin, H. J. (1955). On generalizations of Tchebychef's inequality. *J. Amer. Statist. Assoc. 50*: 923–945.

——— (1964). *Inequalities on Distribution Functions.* Griffin, London.

Greenleaf, Frederick P. (1969). *Invariant Means on Topological Groups and their Applications.* Van Nostrand, New York.

ter Haar, D. (1954). *Elements of Statistical Mechanics.* Rinehart, New York.

Halmos, Paul R. (1953). *Lectures on Ergodic Theory.* Math. Soc. of Japan, Tokyo.

Hausdorff, Felix (1914). *Grundzüge der Mengenlehre*, 1st ed. Von Veit, Leipzig, repr. Chelsea, New York, 1949.

Hewitt, Edwin, and Leonard Jimmie Savage (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc. 80*: 470–501.

Heyde, Christopher C., and Eugene Seneta (1977). *I. J. Bienaymé: Statistical Theory Anticipated.* Springer, New York.

Hopf, Eberhard (1937). *Ergodentheorie.* Springer, Berlin.

——— (1954). The general temporally discrete Markoff process. *J. Rat. Mech. Anal.* (Indiana Univ.) *3*: 13–45.

Ionescu Tulcea, Cassius (1949–1950). Mesures dans les espaces produits. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (Ser. 8) *7*: 208–211.

Jacobs, Konrad (1963). *Lecture Notes on Ergodic Theory.* Aarhus Universitet, Matematisk Institut.

Jessen, Borge (1939). Abstrakt Maal- og Integralteori 4. *Mat. Tidsskrift (B) 1939*, pp. 7–21.

Jordan, Károly (1972). *Chapters on the Classical Calculus of Probability.* Akadémiai Kiadó, Budapest.

Kakutani, Shizuo (1943). Notes on infinite product measures, *I. Proc. Imp. Acad. Tokyo* (became *Japan Academy Proceedings*) *19*: 148–151.

Khinchin, A. Ya. (1933). Zu Birkhoffs Lösung des Ergodenproblems. *Math. Ann. 107*: 485–488.

Kolmogorov, Andrei Nikolaevich (1929a). Bemerkungen zu meiner Arbeit "Über die Summen zufälliger Grössen." *Math. Ann. 102*: 484–488.

Russian). In *Coll. Works, Math. Sect.* (Communist Acad., Sect. Nat. Exact Sci.) *1*, 8–21. Izd. Komm. Akad., Moscow. Repr. in Kolmogorov (1986), pp. 48–58.

——— (1930). Sur la loi forte des grandes nombres. *Comptes Rendus Acad. Sci. Paris 191*: 910–912.

——— (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Math., Springer, Berlin. English transl. *Foundations of the Theory of Probability*, Chelsea, New York, 1956.

——— (1986). *Probability Theory and Mathematical Statistics* [in Russian; selected works], ed. Yu. V. Prohorov. Nauka, Moscow.

Łomnicki, Z., and Stanislaw Ulam (1934). Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités: I. Variables indépendantes. *Fund. Math. 23*: 237–278.

*Markov, Andrei Andreevich (1899). The law of large numbers and the method of least squares (in Russian). *Izv. Fiz.-Mat. Obshch. Kazan Univ.* (Ser. 2) *8*: 110–128.

Móri, T. F., and G. J. Székely (1983). On the Erdös-Rényi generalization of the Borel-Cantelli lemma. *Stud. Sci. Math. Hungar. 18*: 173–182.

von Neumann, Johann (1929). Zur allgemeinen Theorie des Masses. *Fund. Math. 13*: 73–116.

——— (1935). *Functional Operators*. Mimeographed lecture notes. Institute for Advanced Study, Princeton, N.J. Published in *Ann. Math. Studies* no. 21, *Functional Operators*, vol. I, *Measures and Integrals*. Princeton University Press, 1950.

Notices, Amer. Math. Soc. 28 (1981, p. 84; unsigned). 1980 Wolf Prize.

Ornstein, Donald S., D. J. Rudolph, and B. Weiss (1982). Equivalence of measure preserving transformations. *Amer. Math. Soc. Memoirs 262*.

Plancherel, Michel (1912). Sur l'incompatibilité de l'hypothèse ergodique et des équations d'Hamilton. *Archives sci. phys. nat.* (Ser. 4) *33*: 254–255.

——— (1913). Beweis der Unmöglichkeit ergodischer mechanischer Systeme. *Ann. Phys.* (Ser. 4) *42*: 1061–1063.

Révész, Pal (1968). *The Laws of Large Numbers*. Academic Press, New York.

Rosenthal, Arthur (1913). Beweis der Unmöglichkeit ergodischer Gassysteme. *Ann. Phys.* (Ser. 4) *42*: 796–806.

Savage, I. Richard (1961). Probability inequalities of the Tchebycheff type. *J. Research Nat. Bur. Standards 65B*, pp. 211–222.

Seneta, Eugene (1992). On the history of the strong law of large numbers and Boole's inequality. *Historia Mathematica 19*: 24–39.

Sheynin, O. B. (1968). On the early history of the law of large numbers. *Biometrika 55*, pp. 459–467.

Shiryayev, A. N. (1989). Kolmogorov: Life and creative activities. *Ann. Probab. 17*: 866–944.

The Times [London, unsigned] (26 October 1987). Andrei Nikolaevich Kolmogorov: 1903–1987. Repr. in *Inst. Math. Statist. Bulletin 16*: 324–325.

Ulam, Stanisław (1932). Zum Massbegriffe in Produkträumen. *Proc. International Cong. of Mathematicians* (Zürich), *2*, pp. 118–119.

Yosida, Kosaku, and Shizuo Kakutani (1939). Birkhoff's ergodic theorem and the maximal ergodic theorem. *Proc. Imp. Acad.* (Tokyo) *15*: 165–168.

Youschkevitch [Yushkevich], Alexander A. (1974). Markov, Andrei Andreevich. *Dictionary of Scientific Biography, 9*, pp. 124–130.

Youschkevitch, A. P. (1971). Chebyshev, Pafnuti Lvovich. *Dictionary of Scientific Biography, 3*, pp. 222–232.

第9章　依L收敛与中心极限定理

设 X_j 是独立同分布随机变量，均值为0，方差为1，$S_n := X_1 + \cdots + X_n$. 本章将证明概率论的主要定理之一——中心极限定理，并给出了 $S_n/n^{1/2}$ 的依L收敛，从这个意义上讲，对于每个实数 x，有

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{n}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\mathrm{d}t.$$

均值为0、方差为1的假设并没有真正限制了定理的一般性，因为对于任意均值为 m、方差为有限正数 σ^2 的独立同分布随机变量 Y_i，可以通过变换 $X_i = (Y - m_i)/\sigma$ 就可以得到满足上述条件的 X_i，即可以应用此定理，从而得出 $(Y_1 + \cdots + Y_n - nm)/(\sigma n^{1/2})$ 依分布收敛到上式右端给定的函数，即“标准正态”分布．因此应用中心极限定理，对随机变量(可以假定它们是非常数，独立同分布的)分布的唯一限制是 $EY_j^2 < \infty$．那么，无论 Y_j 的最初分布形式怎样(离散的或连续的，对称的或非对称的，等等)，它们部分和的极限分布都有相同的标准形式，这是一个值得注意的事实．随机变量是同分布的这个要求可以减弱(林德伯格定理，见9.6节)，对于多维随机变量，中心极限定理也有不同的形式，其叙述及证明将在9.5节给出．(另外，随机变量 $S_n/n^{1/2}$ 不依概率收敛——事实上，如果 n 远大于 k，S_n 与 S_k 几乎是独立的)．

9.1　分布函数和密度函数

定义　$\mathbb{R}$ (或任意可分度量空间)上的法则就是定义在博雷尔 σ-代数上的任意概率测度．给定某概率空间$(\Omega, \mathcal{S}, P)$上的一个实值随机变量 X，且$\mathcal{L}(X) := p \circ X^{-1} = Q$，$X$ 或 Q 的(累积)**分布函数**(distribution function)是$\mathbb{R}$上由

282

$$F(x) := Q((-\infty, x]) = P(X \leqslant x)$$

给出的函数 F.

这类函数可以特征化如下．

9.1.1 定理　定义在$\mathbb{R}$上的函数 F 为分布函数，当且仅当它非减、右连续，并且满足 $\lim_{x\to-\infty} F(x) = 0$，$\lim_{x\to+\infty} F(X) = 1$. 在$\mathbb{R}$上对于给定的分布函数 F 存在唯一的法则 L.

证明　若 F 是一个分布函数，那么它具有概率测度所具有的性质(非负性、可数可加性和归一性)．反之，如果 F 具有这样的性质，根据定理3.2.6，存在唯一的法则与分布函数 F 相对应．　□

例：函数 $F(x) = \begin{cases} 0, & 若\ x<0 \\ 1, & 若\ x\geqslant 0 \end{cases}$ 是法则 δ_0 以及满足 $p(X=0)=1$ 的任意随机变量 X 的分布函数，如果定义 F 在0点的值为除1以外的任意值，例如0，它将不再是一个右连续函数，因此也就不是一个分布函数．

对于[0，1]上的均匀分布(法则)，函数

$$F(x) = \begin{cases} 0, & 若\ x \leqslant 0 \\ x, & 若\ 0 < x \leqslant 1 \\ 1, & 若\ x > 1 \end{cases}$$

是分布函数.

在任意空间(如$\mathbb{R}$)上对于给定的法则，为了找到与其相对应的随机变量，我们可以把这个空间当作概率空间并且把与其相对应的随机变量当作概率空间上的随机变量. 有时在特定的概率空间上定义随机变量是有用的，如下所示，把$(\Omega, \mathcal{S}, P)$当作具有博雷尔 σ-代数和勒贝格测度的(0, 1)开区间，对$\mathbb{R}$上的任意分布函数 F，令

$$X(t) := X_F(t) := \inf\{x : F(x) \geqslant t\}, \qquad 0 < t < 1.$$

在这个例子中，注意到 $F(x) = 1_{\{x \geqslant 0\}}$，$X_F(t) = 0(0 < t < 1)$，如果 F 是$[0, 1]$上均匀分布的分布函数，则对 $0 < t < 1$，$X_F(t) = t$.

283

9.1.2 命题 对任意分布函数 F，X_F 是与分布函数 F 相对应的随机变量.

证明 显然，如果 $t \leqslant F(x)$，则有 $X(t) \leqslant x$. 反之，如果 $X(t) \leqslant x$，那么由于 F 是非减的，则对任意的 $y > x$，有 $F(y) \geqslant t$，且由右连续性可得 $F(x) \geqslant t$. 因为 X 也是非减的，所以它是可测的，从而

$$P\{t : X(t) \leqslant x\} = P\{t : t \leqslant F(x)\} = F(x).$$ □

下面将给出独立随机变量在给定的法则上的运算.

定义 给定$\mathbb{R}^k$上的两个有限符号测度 μ 和 ν，它们在$\mathbb{R}^k$中每一博雷尔集上的卷积(convolution)定义为

$$(\mu * \nu)(A) := \int_{\mathbb{R}^k} \nu(A - x)\,\mathrm{d}\mu(x), \text{其中 } A - x := \{z - x : z \in A\}.$$

9.1.3 定理 如果 X 和 Y 是$\mathbb{R}^k$上的独立有值随机变量，$\mathcal{L}(X) = \mu$，$\mathcal{L}(Y) = \nu$，则$\mathcal{L}(X + Y) = \mu * \nu$. 因此 $\mu * \nu$ 是概率测度.

证明 由独立性可得出，$\mathbb{R}^{2k}$上的$\mathcal{L}(\langle X, Y\rangle)$是积测度 $\mu \times \nu$. 因为 $X + Y \in A$ 当且仅当 $Y \in A - X$，因此 $1_A(x + y) = 1_{A-x}(y)$，由 Tonelli-Fubini 定理即可得出结论. □

9.1.4 定理 有限博雷尔测度的卷积满足交换律和结合律，

$$\mu * \nu \equiv \nu * \mu, (\mu * \nu) * \rho \equiv \mu * (\nu * \rho).$$

证明 给定有限博雷尔测度 μ，ν，ρ 和任意博雷尔集 A，有

$$((\mu * \nu) * \rho)(A) = (\mu \times \nu \times \rho)\{\langle x, y, z\rangle : x + y + z \in A\},$$

改变运算顺序，上式仍成立. □

$\mathbb{R}$上的法则 P 称为有密度 f，当且仅当 P 关于勒贝格测度 λ 绝对连续并且有 Radon-Nikodym 导数 $\mathrm{d}p/\mathrm{d}\lambda = f$. 换句话说，对所有的博雷尔集 A，$P(A) = \int_A f(x)\,\mathrm{d}x$. 那么，如果 F 是 P 的分布函数，则有

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, x \in \mathbb{R}$$

284

且对于 λ-几乎所有的 x，$F'(x)$都存在且等于 $f(x)$(根据定理 7.2.1). $\mathbb{R}$上的函数 f 是概率密度函数，当且仅当对于勒贝格测度，它是几乎处处非负可测的，且有

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

9.1.5 命题 如果 P 有一个密度函数 f，那么对任意可测的函数 g，$\int g\mathrm{d}p = \int gf\mathrm{d}\lambda$. 等式的一边是有定义的(有限的)，当且仅当另一边是有定义的(有限的).

证明 此证明留作习题11. □

9.1.6 命题 如果 P 和 Q 是定义在$\mathbb{R}$上的两个法则，P 有密度函数f，那么 $P*Q$ 有密度函数$h(x):=\int f(x-y)\mathrm{d}Q(y)$. 如果 Q 有密度函数 g，那么 $h(x)\equiv\int f(x-y)g(y)\mathrm{d}y$.

证明 对于任意博雷尔集 A，根据 Tonelli-Fubini 定理和命题9.1.5可得，

$$\begin{aligned}(P*Q)(A) &= \int P(A-y)\mathrm{d}Q(y) = \iint 1_A(x+y)\mathrm{d}P(x)\mathrm{d}Q(y) \\ &= \iint 1_A(x+y)f(x)\mathrm{d}x\mathrm{d}Q(y) = \iint 1_A(u)f(u-y)\mathrm{d}u\mathrm{d}Q(y) \\ &= \int_A h(u)\mathrm{d}u.\end{aligned}$$

如果 Q 有密度函数 g，h 的形式也如命题中所示. □

例：若 P 是标准指数分布，有密度函数$f(x):=\begin{cases}e^{-x}, & 若\ x\geqslant 0\\ 0, & 若\ x<0\end{cases}$，那么 $P*P$ 有密度函数

$$(f*f)(x) = \begin{cases}\int_0^\infty e^{-(x-y)}1_{\{x-y\geqslant 0\}}e^{-y}\mathrm{d}y = xe^{-x}, & 若\ x\geqslant 0\\ 0, & 若\ x<0.\end{cases}$$

这个例题的推广见习题12.

习题

1. 设 X 是$[2, 6]$上的均匀分布，那么对任意博雷尔集 $A\subset R$，$P(X\in A)=\lambda(A\cap[2, 6])/4$，其中 λ 是勒贝格测度. 试求 X 的分布函数和密度函数.

285

2. 令 k 是4次独立试验中成功的次数，且每次试验成功的概率是0.4，试对于所有实数 x，求出 k 的分布函数.
3. 设 X 和 Y 是独立的且在$[0, 1]$上均服从均匀分布，那么对于任意博雷尔集 $B\subset\mathbb{R}$，$P(X\in B)=P(Y\in B)=\lambda(B\cap[0, 1])$. 试求 $X+Y$ 的分布函数和密度函数.
4. 若 X 和 Y 是独立的且对 $j=1, \cdots, n$，$P(X=j)=P(Y=j)=1/n$，试求 $X+Y$ 的分布函数.
5. 如果 X 是法则为 P 的随机变量，其中对应于 P 的分布函数为 F，证明：$F(x)$在$[0, 1]$上有法则 λ，当且仅当 P 是非原子的，也就是说，对于单点值 x，有 $P(\{x\})=0$.
6. 令 X_n 是均值为0、方差为1的独立实值随机变量序列，Y 是一实值随机变量且 $EY^2<\infty$. 证明：当 $n\to\infty$ 时，$E(X_nY)\to 0$. [提示：利用贝塞尔不等式(5.4.3)和定理8.1.2.]
7. 设 X_1，X_2，…是有相同分布函数 F 的独立随机变量.
   (a)试求 $M_n:=\max(X_1, X_2, \cdots, X_n)$的分布函数.
   (b)如果 $F(x)=\begin{cases}1-e^{-x}, & 若\ x>0\\ 0, & 若\ x\leqslant 0\end{cases}$，证明：当 $n\to\infty$ 时，$M_n-\log n$ 的分布函数是收敛的，并求其极限.
8. 对任一随机变量 X，令其分布函数为 F_X，试按照 F_X 的形式表示出 F_{-X}.
9. 写出下列条件下 F 的取值范围$\{F(x): x\in\mathbb{R}\}$:
   (a)分布函数 F 是连续的.
   (b)分布函数 F 为一般的分布函数.
10. 若 F 是定义于有理数集$\mathbb{Q}$上概率法则 P 的分布函数，那么 $P(\mathbb{Q})=1$. 令 H 是 F 的值域. 下面的结论若成立，则证明之，否则举反例：
    (a)H 总是可数的. [提示：在命题4.2.1中反变换 g.]

(b) H 总有勒贝格测度为 0.

11. 证明命题 9.1.5.

12. Γ 函数(gamma function)定义为 $\Gamma(a) := \int_0^\infty x^{a-1}e^{-x}dx(a>0)$，那么 $f_a(x) := e^{-x}x^{a-1}1_{\{x>0\}}/\Gamma(a)$ 是法则 Γ_a 的密度函数. β 函数(beta function)定义为 $\beta(a, b) := \int_0^1 x^{a-1}(1-x)^{b-1}dx(a>0, b>0)$. 证明：

(i) $\Gamma_a * \Gamma_b = \Gamma_{a+b}$.

(ii) $\beta(a, b) \equiv \Gamma(a) * \Gamma(b)/\Gamma(a+b)$.

[提示：本节的最后一个例子已经证得 $\Gamma_1 * \Gamma_1 = \Gamma_2$. 要求得 $\Gamma_a * \Gamma_b$，作换元 $y = xu$，有 $\int_0^x (x-y)^{a-1}y^{b-1}dy = x^{a+b-1}\beta(a, b)$. 注意对于 Γ_c，标准化常数 $1/\Gamma(c)$ 是唯一的，然后用 $c=a+b$ 来证明(i)和(ii).]

286

13. (a)用归纳法证明 $k! = \Gamma(k+1)$，其中 $k=0, 1, \cdots$.

(b)对于泊松概率 $Q(k, \lambda) := \sum_{j=k}^{\infty} e^{-\lambda}\lambda^j/j!$ 和习题 12 中 Γ 的概率密度函数为 f_a，证明：$Q(k,\lambda) = \int_0^\lambda f_k(x)dx$，其中 $\lambda>0$，$k=1, 2, \cdots$.

(c)对于二项概率分布 $E(k, n, p) = \sum_{j=k}^{n}\binom{n}{j}p^j(1-p)^{n-j}$，证明 $E(k, n, p) = \int_0^p x^{k-1}(1-x)^{n-k}dx/\beta(k, n-k+1)$，其中 $0<p<1$，$k=1, \cdots, n$.

9.2 随机变量的收敛性

本章的大部分及后面的章节中，随机变量是在 $\mathbb{R}$ 上或有限维欧几里得空间 $\mathbb{R}^k$ 上取值的，但其定义和一些结论可以推广到(可分)度量空间.

设 $(S, \mathcal{T})$ 是一个拓扑空间，$(\Omega, \mathcal{A}, P)$ 是一概率空间. 令 $Y_0, Y_1, \cdots$ 是定义在 Ω 上取值于 S 的随机变量. Y 对于由 $\mathcal{T}$ 生成的 S 中博雷尔集的 σ-代数是可测的，这等价于对所有的开集 U，有 $Y^{-1}(U) \in \mathcal{A}$(由定理 4.1.6). 那么 $Y_n \to Y_0$ a. s.，通常意味着对几乎所有的 ω，$Y_n(\omega) \to Y_0$. 如果 (S, d) 是一个度量空间，假定 $Y_n(n\geqslant 1)$ 是(可测的)随机变量并且 $Y_n \to Y_0$ a. s.，则根据定理 4.2.2 可知，Y_0 也是一随机变量(至少对于 P 的完备化). 这在更一般的拓扑空间上是行不通的(命题 4.2.3). 度量空间有一个很好的性质：贝尔 σ-代数和博雷尔 σ-代数是等价的(定理 7.1.1).

如果 (S, d) 是一可分的度量空间(有可数的稠密集)，那么它的拓扑是第二可数的(命题 2.1.4). 因此对于这样的两个空间的笛卡儿积，积拓扑的博雷尔 σ-代数等于各自空间上博雷尔 σ-代数的积 σ-代数(命题 4.1.7). 度量 d 是连续的，则在 S 与其自身的积空间上是可测的，因此 d 对于积 σ-代数是联合可测的. 对取值于 S 的任意两个随机变量 X 和 Y，$d(X, Y)$ 是一个随机变量. 因此下面的定义有意义.

定义 *对于任意可分度量空间 (S, d)，定义在概率空间 $(\Omega, \mathcal{S}, P)$ 取值于 S 的随机变量 Y_n 依概率收敛到 Y_0，当且仅当对于任意 $\varepsilon>0$，当 $n\to\infty$ 时，有 $P\{d(Y_n, Y_0)>\varepsilon\}\to 0$.*

287

对于在任意可分度量空间上取值的随机变量，显然几乎必然收敛蕴涵着依概率收敛，反之则不一定成立. 例如，令 $A(1), A(2), \cdots$ 是任意满足 $P(A(n))\to 0(n\to\infty)$ 的事件，那么 $1_{A(n)}$ 依概率收敛到 0. 根据(博雷尔–坎泰利引理 8.3.4)可得，如果这些事件是独立的，那么 $1_{A(n)}\to 0$ a. s. 当且仅当 $\sum_n P(A(n)) < \infty$. 因此，如果对所有的 n，$P(A(n)) = 1/n$，则此序列依概率收敛到 0，但不是几乎必然收敛. 然而，如果用子序列来讨论，则结论是相反的，从而给出这两种收敛之间的密切

关系.

9.2.1 定理 对任意定义在概率空间$(\Omega, \mathcal{S}, P)$取值于可分度量空间(S, d)的随机变量X_n和X，X_n依概率收敛到X当且仅当对每个子序列$X_{n(k)}$，有一个子序列$X_{n(k(r))} \to X$a. s..

证明 如果X_n依概率收敛到X，则它的任一子序列都收敛到X. 考虑$X_{n(k)}$，如果取$k(r)$使其满足

$$P\{d(X_{n(k(r))}, X) > 1/r\} < 1/r^2, \qquad r = 1,2,\cdots,k(r) \to \infty,$$

那么根据博雷尔–坎泰利引理(8.3.4)，对足够大的r(依赖ω)，有$d(X_{n(k(r))}, X) \leqslant 1/r$a. s.，因此$X_{n(k(r))} \to X$a. s..

反之，如果X_n不依概率收敛到X，那么存在$\varepsilon > 0$和子序列$X_{n(k)}$，使得对所有的k，

$$P\{d(X_{n(k)}, X) > \varepsilon\} \geqslant \varepsilon.$$

那么这个子序列没有子序列几乎必然收敛到X. □

例：设$A(1)$，$A(2)$，…是满足$P(A(n)) \to 0 (n \to \infty)$的任意事件，且$1_{A(n)}$依概率收敛到0，但不必是几乎必然收敛，那么对某子序列$n(k)$，对所有的$k = 1, 2, \cdots$，$P(A(n(k))) < 1/k^2$，且$1_{A(n(k))} \to 0$a. s..

对于任意可测空间$(\Omega, \mathcal{A})$和$(S, \mathcal{B})$，对于给定的σ-代数，令$\mathcal{L}^0(\Omega, \mathcal{A}; S, \mathcal{B})$为从$\Omega$映射到$S$的所有可测函数的集合. 如果$\mu$是$\mathcal{A}$上的测度，令$L^0(\Omega, \mathcal{A}, \mu; S, \mathcal{B})$是$\mathcal{L}^0(\Omega, \mathcal{A}; S, \mathcal{B})$中与$\mu$几乎处处相等的元素的等价类的集合. 如果$S$是可分度量空间，记号中的$\sigma$-代数$\mathcal{B}$被略去，其可理解为博雷尔$\sigma$-代数. 同理，一旦把$\mathcal{A}$和$\mu$具体化，不必要再标明，因此可写为$L^0(\Omega, S)$. 如果$S$没有具体化，则可理解为实线$\mathbb{R}$，此时可记为$L^0(\mu)$或$L^0(\Omega)$.

288

如果某些随机变量被其他与其几乎必然相等的随机变量所代替，它们的几乎必然收敛性或依概率收敛性不变，因此，这些收敛方式既可定义在$\mathcal{L}^0$上也可定义在L^0上.

对任意可分度量空间(S, d)、概率空间$(\Omega, \mathcal{A}, P)$和$X, Y \in \mathcal{L}^0(\Omega, S)$，令

$$\alpha(X,Y) := \inf\{\varepsilon \geqslant 0 : P(d(X,Y) > \varepsilon) \leqslant \varepsilon\}.$$

因此，对某$\varepsilon_k \downarrow \alpha := \alpha(X, Y)$，对所有的$k \geqslant j$，$P(d(X, Y) > \varepsilon_k) \leqslant \varepsilon_k \leqslant \varepsilon_j$. 当$k \to \infty$时，$\{d(X, Y) > \varepsilon_k\}$的示性函数单调递增到$\{d(X, Y) > \alpha\}$，因此由单调收敛性可得，对所有的$j$，$P\{d(X, Y) > \alpha\} \leqslant \varepsilon_j$且$P\{d(X, Y) > \alpha\} \leqslant \alpha$. 换句话说，$\alpha(X, Y)$的下确界是可以达到的.

例：对于两个常值函数$X \equiv a$和$Y \equiv b$，$\alpha(X, Y) = \min(1, |a - b|)$，则对任意事件$B$，$\alpha(1_B, 0) = P(B)$.

9.2.2 定理 在$L^0(\Omega, S)$上，α是一个度量，其度量化依概率收敛，所以$\alpha(X_n, X) \to 0$当且仅当X_n依概率收敛到X.

证明 显然α具有对称性和非负性，且$\alpha(X, Y) = 0$当且仅当$X = Y$a. s.. 对三角不等式，考虑随机变量X，Y和Z，除了一个概率很小的集合外，$d(X, Y)$不超过$\alpha(X, Y)$，都有$d(X, Y) \leqslant \alpha(X, Y)$，同理，对$Y$和$Z$有相同的结论. 因此，在$S$中应用三角不等式，除了一个概率很小的集合外，对大多数$\alpha(X, Y) + \alpha(Y, Z)$，都有

$$d(X,Z) \leqslant d(X,Y) + d(Y,Z) \leqslant \alpha(X,Y) + \alpha(Y,Z).$$

因此$\alpha(X, Z) \leqslant \alpha(X, Y) + \alpha(Y, Z)$，并且$\alpha$是$L^0$的一个度量.

如果X_n依概率收敛到X，那么对每个$m = 1, 2, \cdots$当n(如$n \geqslant n(m)$)足够大时，有$P(d(X_n, X) > 1/m) \leqslant 1/m$. 故当$n \to \infty$时，有$\alpha(X_n, X) \to 0$. 反之，对于$\alpha$，如果$X_n \to X$，那么对

于任意的 $\delta>0$，当 n 足够大时，有 $\alpha(X_n, X)<\delta$. 因此有 $P(d(X_n, X)>\delta)<\delta$，从而可得 X_n 依概率收敛到 X. □

度量 α 称为樊畿(Ky Fan)度量.

由定理 9.2.1 和定理 9.2.2 可知，几乎必然收敛是可度量化的，当且仅当它是依概率收敛的. 如果(Ω, P)是纯原子的，这是成立的，所以存在 $\omega_k\in\Omega$，使得 $\sum_k P\{\omega_k\}=1$，但是通常情况下这是不成立的(见习题 1 ~ 3).

289

9.2.3 定理 如果(S, d)是完备可分的度量空间，$(\Omega, \mathcal{S}, P)$是任意概率空间，那么 $L^0(\Omega, S)$对樊畿度量 α 是完备的.

证明 考虑柯西序列$\{X_n\}$，设子序列$\{X_{n(r)}\}$满足 $\sup_{m\geqslant n(r)}\alpha(X_m, X_{n(r)})\geqslant 1/r^2$，那么对所有的 $s\geqslant r$，$P\{d(X_{n(r)}, X_{n(s)})>1/r^2\}\leqslant 1/r^2$. 因为 $\sum r^{-2}$ 是收敛的，博雷尔-坎泰利引理蕴涵了对足够大的 r，$d(X_{n(r)}, X_{n(r+1)})\leqslant 1/r^2$ 几乎必然成立. 因此对所有的 $s\geqslant r$，$d(X_{n(r)}, X_{n(s)})\leqslant\sum_{j\geqslant r}j^{-2}<1/(r-1)$. 那么$\{X_{n(r)}(\omega)\}_{r\geqslant1}$是柯西序列且收敛到某个 $Y(\omega)$，因为 S 是完备的.(假定 $X_{n(r)}$ 不收敛，任意定义 Y.)那么 $\alpha(X_{n(r)}, Y)\to0$. 在任意度量空间中，柯西序列和它的收敛子列收敛到相同的极限，因此 $\alpha(X_m, Y)\to0$. □

下面的定理给出了几乎必然收敛的“柯西准则”. 尽管它并不很新奇，但是由于几乎必然收敛不是可度量的，因此它的证明颇有意义，例如，对$[0, 1]$上的勒贝格测度.

9.2.4 引理 令(S, d)是一完备可分的度量空间，$(\Omega, \mathcal{A}, P)$是一概率空间. 令 $Y_1, Y_2, \cdots$是从 Ω 映射到 S 的随机变量，且对任意 $\varepsilon>0$，满足

$$P\left\{\sup_{k\geqslant n}d(Y_n, Y_k)\geqslant\varepsilon\right\}\to0, n\to\infty.$$

那么对某个随机变量 Y，Y_n 几乎必然收敛于 Y.

证明 令 $A_{mn}:=\{\sup\{d(Y_j, Y_k): j, k\geqslant n\}\leqslant 1/m\}$. 对每个 m，A_{mn}随 n 递增，且 A_{mn}是可测的. 对 $\varepsilon=1/(2m)$，应用三角不等式 $d(Y_j, Y_k)\leqslant d(Y_j, Y_n)+d(Y_n, Y_k)$得，$P(A_{mn})\uparrow1(n\to\infty)$. 令 $A_m:=\bigcup_n A_{mn}$，那么对所有的 m，有 $P(A_m)=1$，$P\left(\bigcap_m A_m\right)=1$. 对任意 $\omega\in\bigcap_m A_m$，$Y_n(\omega)$几乎必然收敛到某个 $Y(\omega)$. 对 $\omega\notin\bigcap_m A_m$，定义 $Y(\omega)$是 S 中任意固定的点，则由定理 4.2.2(和引理 4.2.4)知 $Y(\cdot)$是可测的. 证毕. □

习题

1. 对 $n=1, 2, \cdots$和 $2^{n-1}\leqslant j<2^n$，当 $j/2^{n-1}-1\leqslant x\leqslant(j+1)/2^{n-1}-1$ 时，令 $f_j(x)=1$；在其他情况时，令 $f_j=0$. 证明：对$[0, 1]$上的一致(勒贝格)概率测度，当 $j\to\infty$ 时，f_j 依概率收敛到 0 但不是几乎必然收敛. 并求一个子序列 $f_j(r)(r=1, 2, \cdots)$，使得当 $r\to\infty$ 时，$f_j(r)$几乎必然收敛到 0.

290

2. 根据 2.1 节习题 10 的定义证明：在 L^0 上几乎必然收敛是 L-收敛，依概率收敛是 L^*-收敛. 假定 2.1 节习题 10 和本节习题 1 的结论成立，证明：在 L^0 上不存在拓扑，使得对具有勒贝格测度$[0, 1]$上几乎必然收敛的随机变量对此拓扑是收敛的.
3. 设$(X, 2^X, P)$是一个概率空间，X 是一可数集，2^X 是一个幂集(X 的所有子集构成的 σ-代数). 证明：在 X 上几乎必然收敛与依概率收敛是等价的.
4. 设 $f(x):=x/(1+x)$，$\tau(X, Y):=\int f(d(X, Y))\mathrm{d}P$. 证明：$\tau$ 是 L^0 上的一个度量并且它的度量依概率收敛.

[提示：见命题2.4.3和命题2.4.4.]

5. 对定义于$L^0(\Omega, S)$、取值于度量空间(S, d)的X，Y，令$\varphi(X, Y):=\inf\{\varepsilon+P\{\omega: d(X, Y)>\varepsilon\}: \varepsilon>0\}$，设$\alpha$是樊畿度量，$\tau$的定义和前面习题中的一样．试求有限常数$C_1$，$C_2$，$C_3$，$C_4$，使得对所有的$(S, d)$和在$S$中取值的随机变量$X$，$Y$，以及$\alpha:=\alpha(X, Y)$，$\varphi:=\varphi(X, Y)$，$\tau:=\tau(X, Y)$，有$\varphi\leqslant C_1\alpha$，$\alpha\leqslant C_2\varphi$，$\tau\leqslant C_3\alpha$，$\alpha\leqslant C_4\tau^{1/2}$.
6. 证明：在$L^0(\lambda)$中，其中λ是$[0, 1]$上的勒贝格测度，对于习题5中的所有随机变量X，Y，不存在$C<\infty$，使得$\alpha\leqslant C\tau$.
7. 设P，Q是同一σ-代数$\mathcal{A}$上的两个概率测度，它们是等价的意味着，$P(A)=0$当且仅当对所有的$A\in\mathcal{A}$，$Q(A)=0$. 证明：P依概率收敛和Q依概率收敛是等价的．
8. 对习题7中定义的P，Q，令α_P，α_Q是相对应的樊畿度量．试举一例：P和Q等价且不存在$C<\infty$，满足对任意$\mathcal{A}$可测的实值随机变量有$\alpha_P(X, Y)\leqslant C\alpha_Q(X, Y)$.

9.3 依分布收敛

令$(S, \mathcal{T})$是一拓扑空间，P_n是一个法则序列，即由$\mathcal{T}$产生的博雷尔σ-代数$\mathcal{B}$上的概率测度．[291] 为定义P_n收敛到P_0，一种方法是：对任意符号测度μ，有若尔当分解$\mu=\mu^+-\mu^-$（定理5.6.1），及全变差测度$|\mu|:=\mu^++\mu^-\geqslant 0$. 我们称$P_n$依全变差收敛到$P_0$当且仅当$n\to\infty$时，$|P_n-P_0|(S)\to 0$，或等价地

$$\sup_{A\in\mathcal{B}}|(P_n-P_0)(A)|\to 0.$$

然而依全变差收敛对大多数应用要求太强，例如，令$x(n)$是S中收敛到x的点列，回忆由$\delta_y(A):=1_A(y)$（$\forall y\in S$）定义的点质量．假设单元素集$\{x\}$是一个博雷尔集（这是没有疑义的，例如，如果S是可度量化的或是豪斯多夫的）．除了对足够大的n有$x(n)=x$外，$\delta_{x(n)}$不依全变差收敛到δ_x. 事实上，对每个博雷尔集$A\subset\mathbb{R}$，或者对每个开集或闭集（考虑$A=(-\infty, 0]$或$(0, \infty)$），$\delta_{1/n}(A)$收敛到$\delta_0(A)$不成立．P_n依全变差收敛到δ_0当且仅当$P_n(\{0\})\to 1$. 下面的定义给出了与S中的收敛性相关的更好的收敛性，使得无论何时$x(n)\to x$，都有$\delta_{x(n)}$收敛到δ_x.

定义 设$C_b(S)$是S中所有有界连续实值函数的集合．我们称法则P_n收敛到法则P，记作$P_n\underset{\mathcal{L}}{\to}P$，或仅记作$P_n\to P$，当且仅当对每个$f\in C_b(S)$，当$n\to\infty$时，$\int f\mathrm{d}P_n\to\int f\mathrm{d}P$.

注意到，任意$f\in C_b(S)$是有界可测的，且对任意法则是可积的．如果$x(n)\to x$，那么$\delta_{x(n)}\to\delta_x$. 法则的收敛是指对拓扑收敛的，特别地对一个积拓扑、逐点收敛拓扑，把它限定到$C_b(S)$上所有实值函数集的子集上．这就可以得出下面的命题．

9.3.1 命题 如果法则P_n和P使得对每个子序列$P_{n(k)}$都存在一个子序列$P_{n(k(r))}\underset{\mathcal{L}}{\to}P$，那么$P_n\underset{\mathcal{L}}{\to}P$.

证明 如果结论不成立，即存在某$f\in C_b$，$\int f\mathrm{d}P_n\not\to\int f\mathrm{d}P$，那么对任意$\varepsilon>0$和序列$n(k)$，对所有的$k$，$\left|\int f\mathrm{d}P_{n(k)}-\int f\mathrm{d}P\right|>\varepsilon$，这与$P_{n(k(r))}\underset{\mathcal{L}}{\to}P$相矛盾．□

下面将给出度量空间上法则收敛的一些结论及其证明．

9.3.2 引理 如果(S, d)是一度量空间，P和Q是S上的两个法则，且对所有的$f\in C_b(S)$，[292] $\int f\mathrm{d}P=\int f\mathrm{d}Q$，那么$P=Q$.

证明 令U是S的任意开子集，其补集为F，考虑距离$d(x, F)$，对$n=1, 2, \cdots$令$f_n(x):=\min(1,$

$nd(x, F))$，(见图 9-3A)，那么 $f_n \in C_b(S)$，且当 $n\to\infty$ 时，$f_n \uparrow 1_U$. 因此由单调收敛性可得，$P(U)=Q(U)$，进而有 $P(F)=Q(F)$. 那么，由定理 7.1.3(闭正则性)可得 $P=Q$. □

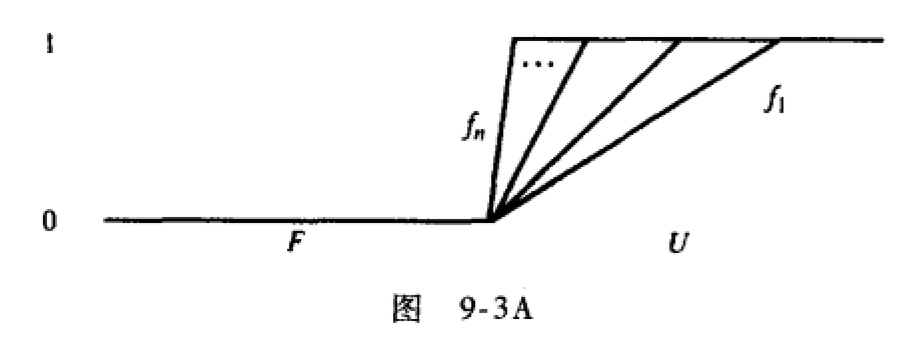


图 9-3A

下面给出一个收敛的法则序列有唯一的极限.

拓扑空间 S 上的法则集$\mathcal{P}$是一致紧的(uniformly tight)，当且仅当对任意 $\varepsilon>0$，存在一紧集 $K\subset S$，使得对所有的 $P\in\mathcal{P}$，$P(K)>1-\varepsilon$. 因此法则 P 是紧的(见 7.1 节的定义)，当且仅当 $\{P\}$ 是一致紧的.

例：$\mathbb{R}$ 上的法则序列 $\{\delta_n\}_{n\geq 1}$不是一致紧的.

下面的事实仅证明了$\mathbb{R}^k$ 上情形. 事实上，对任意度量空间该事实都成立(定理 11.5.4 将在完备可分度量空间上给出其证明).

9.3.3 定理 *设 $\{P_n\}$ 是$\mathbb{R}^k$ 上一致紧的法则序列，那么对某个法则 P，存在子序列 $P_{n(j)}\to P$.*

证明 对每个紧集 $K\subset\mathbb{R}^k$，K 上的所有连续实值函数是有界的，因此，$C_b(K)$ 与 K 上的所有连续实值函数构成的空间 $C(K)$ 是等价的. 由魏尔斯特拉斯逼近定理(推论 2.4.12)可得，对于上确界距离 $s(f, g):=\sup_{x\in K}|f(x)-g(x)|$，($k$ 元)多项式在 $C(K)$ 中是稠密的. 故对有理系数多项式而言，$C(K)$ 是可分的. 设 D 为 $C(K)$ 中的可数稠密集，那么$\left\{\left\{\int f dP_n\right\}_{f\in D}\right\}_{n\geq 1}$是紧区间的可数笛卡儿积 $L:=\prod_{f\in D}[\inf f, \sup f]$ 中的一个序列. 现在具有积拓扑的 L 是紧的(吉洪诺夫定理 2.2.8)且是可度量化的(命题 2.4.4)，因此(由定理 2.3.1)存在一个子序列 $P_{m(j)}$，使得对所有的 $f\in D$，$\int f dP_{m(j)}$ 都收敛. 如果 $f\in C(K)$，$\varepsilon>0$，取 $g\in D$，满足 $s(f, g)<\varepsilon$. 那么对任意的 j，

[293]

$$\left|\int f dP_{m(j)} - \int g dP_{m(j)}\right| < \varepsilon,$$

所以

$$\limsup_{j\to\infty}\int f dP_{m(j)} - \liminf_{j\to\infty}\int f dP_{m(j)} \leq 2\varepsilon.$$

令 $\varepsilon\downarrow 0$，对所有的 $f\in C(K)$，$\int f dP_{m(j)}$ 都收敛.

对每个 $r=1, 2, \cdots$ 取紧集 K_r 使得对所有的 n，都有 $P_n(K_r)>1-1/r$ 成立. 令 $K(r):=K_r$. 对任意的 r，证明的前半部分可应用于 $K=K_r$. 问题变为寻求一个子序列，使得它对所有的 r 都收敛. 过程如下：对每个 r，设 $D(r)$ 在 $C(K_r)$ 中是可数稠密集，则$\left\{\left\{\int f dP_n\right\}_{f\in D(r), r\geq 1}\right\}_{n\geq 1}$是可数积$\prod_{r\geq 1}\prod_{f\in D(r)}[\inf f, \sup f]$ 中的一个序列. 和前面一样，这个序列有一个收敛的子序列，记为 $P_{n(j)}$，且对每个 $f\in C(K_r)$，$\int_{K(r)} f dP_{n(j)}$ 收敛，从而对每个 $\int_{K(r)} f dP_{n(j)}$，$f\in C_b(\mathbb{R}^k)$ 也收敛.

每个不同于$\int f\mathrm{d}P_{n(j)}$的积分不超过 $\sup|f|/r$，当 $r\to\infty$ 时，$\sup|f|/r$ 趋于零．因而对每个 $f\in C_b(\mathbb{R}^k)$，当 $j\to\infty$ 时，$\int f\mathrm{d}P_{n(j)}$ 收敛，且极限记为 $L(f)$.

为了对 L 应用斯通–丹尼尔定理(4.5.2)，注意到 $C_b(\mathbb{R}^k)$ 是斯通向量格，L 是线性的，且当 $f\geqslant 0$ 时，$L(f)\geqslant 0$. 假设 $f_n\downarrow 0$(逐点)，设 $M:=\sup_x f_1(x)$，则对所有的 n 和 x，有 $0\leqslant f_n(x)\leqslant M$. 给定 $\varepsilon>0$，取 r 足够大，使得 $1/r<\varepsilon/(2M)$，由迪尼定理(2.4.10)，在 K_r 上一致地有 $f_n\downarrow 0$. 故存在某个 J，当 $n\geqslant J$ 时，对所有的 $x\in K_r$，$f_n(x)\leqslant\varepsilon/2$. 从而对所有的 m，$\int f_n\mathrm{d}P_m\leqslant\varepsilon$，即 $L(f_n)\leqslant\varepsilon$. 设 $\varepsilon\downarrow 0$，有 $L(f_n)\downarrow 0$. 由斯通–丹尼尔定理(4.5.2)，存在$\mathbb{R}^k$ 上的非负测度 P，使得对所有的 $f\in C_b(\mathbb{R}^k)$，$L(f)=\int f\mathrm{d}P$. 取 $f=1$，证明 P 是一个概率测度，故 $P_{n(j)}\to P$. □

事实上，反过来的结论也是很有用的．

9.3.4 命题 $\mathbb{R}^k$ 上任意一个收敛的法则序列 $P_n\to P$ 是一致紧的．

证明 给定 $\varepsilon>0$，取 $M<\infty$，使得 $P(|x|>M)<\varepsilon$. 设 f 是一个连续实值函数，使得 $|x|\leqslant M$ 时，$f(x)=0$，$|x|\geqslant 2M$ 时，$f(x)=1$，且 $0\leqslant f\leqslant 1$，例如，

294

$$f(x):=\max(0,\min(1,|x|/M-1)).$$

则当 n 足够大(如 $n>r$)时，有 $P(|x|>2M)\leqslant\int f\mathrm{d}P_n<\varepsilon$，对每个 $n\leqslant r$，取 M_n，使得 $P(|x|>M_n)<\varepsilon$. 令 $J:=\max(2M,M_1,\cdots,M_r)$，则对所有的 n，$P(|x|>J)<\varepsilon$. □

取值于拓扑空间 S 的随机变量 X_n 称为依法则收敛或者依分布收敛到随机变量 X，当且仅当 $\mathcal{L}(X_n)\to\mathcal{L}(X)$. 例如，若 X_n 是非常量的独立同分布(i. i. d.)随机变量，它们依法则收敛但不依概率收敛(也不是几乎处处收敛)．对于法则的收敛性，尽管 X_n 通常是定义在相同的空间上，但也可定义在不同的概率空间上．在某些情况下，由 $X_n\underset{\mathcal{L}}{\to}X$ 可以得到$\mathcal{L}(X_n)\underset{\mathcal{L}}{\to}\mathcal{L}(X)$，因而在→下的$\mathcal{L}$可以被省略．

依概率收敛蕴含着依法则收敛．

9.3.5 命题 若(S,d)是一个可分度量空间，X_n 是一个取值于 S 的随机变量，满足 X_n 依概率收敛到 X，则$\mathcal{L}(X_n)\to\mathcal{L}(X)$.

证明 由定理9.2.1知，对任意的序列 $X_{n(k)}$，都存在子序列 $X_{n(k(r))}\to X$ a. s.. 设 $f\in C_b(S)$，由控制收敛定理，$\int f(X_{n(k(r))})\mathrm{d}P\to\int f(X)\mathrm{d}P$，从而由命题9.3.1知，结论成立． □

尽管依法则收敛的随机变量不一定依概率收敛，但是存在命题9.3.5的一个逆命题：如果度量空间上的法则 P_n 收敛于 P_0，则在概率空间(例如，[0, 1])上的勒贝格测度)上，存在随机变量 X_n，使得对所有的 n，$\mathcal{L}(X_n)=P_n$，$X_n\to X_0$ a. s.. 这将在11.7节给出证明．

在实直线上，法则收敛蕴涵分布函数收敛．

9.3.6 定理 在$\mathbb{R}$上，若法则 $P_n\to P_0$，对每个 n，F_n 是 P_n 的分布函数，则对所有的 t，$F_n(t)\to F_0(t)$，其中 F_0 是 t 的连续函数．

证明 对任意的 t，$x\in\mathbb{R}$，且 $\delta>0$，设

$$f_{t,\delta}(x):=\min(1,\max(0,(t-x)/\delta))\qquad(\text{见图 9-3B}).$$

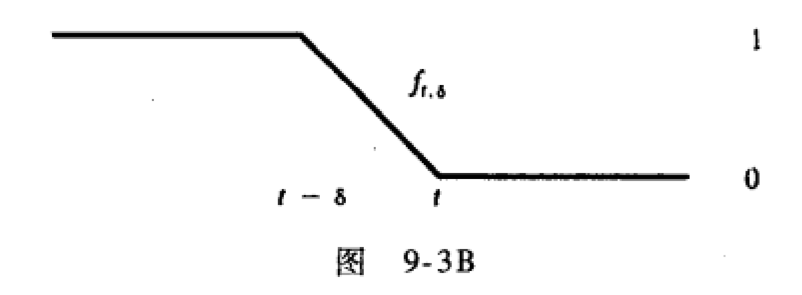


图 9-3B

则当 $x \geqslant t$ 时，$f_{t,\delta}(x)=0$；当 $x \leqslant t-\delta$ 时，$f_{t,\delta}(x)=1$，$0 \leqslant f_{t,\delta} \leqslant 1$，且 $f_{t,\delta} \in C_b(\mathbb{R})$，因此当 $n \to \infty$ 时，有

$$F_n(t) \geqslant \int f_{t,\delta} dP_n \to \int f_{t,\delta} dP_0 \geqslant F_0(t-\delta),$$

$$F_n(t) \leqslant \int f_{t+\delta,\delta} dP_n \to \int f_{t+\delta,\delta} dP_0 \leqslant F_0(t+\delta).$$

因此$\liminf_{n\to\infty} F_n(t) \geqslant F_0(t-\delta)$和$\limsup_{n\to\infty} F_n(t) \leqslant F_0(t+\delta)$. 令 $\delta \downarrow 0$，那么由 F_0 对 t 的连续性可得结论. □

例：$\delta_{1/n} \to \delta_0$，但是其所对应的分布函数不收敛到 0. 这就表明了为什么在定理 9.3.6 中对 t 的连续性是必需的.

通常情况下，一个映射是指一个函数关系，这种函数在一般空间上是连续函数. 连续映射总是保持法则的收敛性.

9.3.7 定理（连续映射定理） *令$(X, \mathcal{T})$和$(Y, \mathcal{U})$是任意两个拓扑空间，G 是从 X 映射到 Y 的连续函数. 令 P_n 是 X 上满足 $P_n \to P_0$ 的法则，那么在 Y 上像法则收敛：$P_n \circ G^{-1} \to P_0 \circ G^{-1}$.*

证明 只要注意到对 Y 上任一有界连续实值函数 f，在 X 上，$f \circ G$ 即是此类函数，运用像测度定理(4.1.11)即得结论. □

习题

1. 设 P_n 是整数集$\mathbb{Z}$上的法则且有 $P_n \underset{\mathcal{L}}{\to} P_0$，证明：$P_n$ 依全变差收敛到 P_0.
2. 设(S, d)是一度量空间，T 是有全变差距离的所有点质量 $\delta_x(x \in S)$的集合，证明：T 上的拓朴是离散的（并且是不依赖于 d 的）.
3. 当 x 取何值时 $\delta_{-1/n}$的分布函数收敛到 δ_0？
4. 设 P 是$\mathbb{R}^2$ 的法则且平行于坐标轴的直线的概率是 0，即用符号表示为：对每一个常数 c，$P(x=c)=P(y=c)=0$. 设法则 P_n 收敛到 P. 证明：对所有的实数 a，b，当 $n \to \infty$ 时，$(P_n-P)((-\infty, a] \times (-\infty, b]) \to 0$.
5. 对任意 $a<b<c<d$，设$f_{a,b,c,d}$是在$(-\infty, a]$和$[d, \infty)$上为 0、在$[b, c]$上为 1 的一个分布函数，其图形在$[a, b]$和$[c, d]$上是一条线段，使得 f 是分段线性函数. 设 P 和 Q 是$\mathbb{R}$上的两个法则且对所有的 $a<b<c<d$，$\int f_{a,b,c,d} \mathrm{d}(P-Q)=0$，证明：$P=Q$.
6. 设 X_n 是取值于可分度量空间 S 的随机变量，且 X_n 依法则收敛到 c，其中 c 为 S 中的点. 证明：X_n 依概率收敛到 c.
7. 设 F 是$[0, \infty)$的示性函数，试举一例：在$\mathbb{R}$上法则序列 $P_n \to P_0$ 但 $P_n \circ F^{-1} \not\to P_0 \circ F^{-1}$.
8. 设 $f_n(x)=\begin{cases}2, & 若(2k-1)/2^n \leqslant x<2k/2^n(k=1, \cdots, 2^{n-1}) \\ 0, & 其他\end{cases}$，法则 P_n 的密度函数为 f_n（关于$[0, 1]$上的勒贝格测度 λ）.

   (a) 证明：$P_n \to \lambda$.

295
≀
296

(b)证明：对[0, 1]上每个有界可测的f, $\int f\mathrm{d}P_n \to \int f\mathrm{d}\lambda$.[提示：$f_n - 1$关于$\lambda$是正交的，并利用(5.4.3).]

(c)对所有的n，试求全变差距离$\|P_n - \lambda\|$.

9. 对于下面$\mathbb{R}$上关于勒贝格测度密度函数为f_n的法则序列P_n，哪个是一致紧的？

(a)$f_n = 1_{[0,n]}/n$.

(b)$f_n(x) = ne^{-nx}1_{[0,\infty)}(x)$.

(c)$f_n(x) = e^{-x/n}1_{[0,\infty)}(x)/n$.

10. 每一个正整数n可用其素因子分解表示出来，如对非负整数n_2，n_3，…，$n = 2^{n_2}3^{n_3}5^{n_5}7^{n_7}11^{n_{11}}\cdots$，令$m(n) := n_2 + n_3 + \cdots$是一个收敛级数．对每个素数$p = 2$，3；5，…在$p$处，令法则$P_n$取值为$n_p/m(n)$.

(a)证明：$\{P_n\}_{n\geq 1}$不是一致紧的．

(b)证明：法则序列$\{P_n\}$有收敛子序列$\{P_{n(k)}\}$.

297 (c)求一收敛子序列$P_{n(k)}$，使得对每一素数p存在满足$n(k)_p > 0$的k.

9.4 特征函数

和在$\mathbb{R}^1$中一样，如果$\mathbb{R}^k$上的函数$f\geq 0$，f关于勒贝格测度λ^k($\mathbb{R}$上勒贝格测度λ的k次积测度，如同定理4.4.6)是可测的，且$\int f\mathrm{d}\lambda^k = 1$，则$f$称为概率密度(probability density)函数．那么$\mathbb{R}^k$上的法则P可由下式给出：对所有的博雷尔(或勒贝格)可测度集A，$P(A) := \int_A f\mathrm{d}\lambda^k$．从而$f$是Radon-Nikodym导数$\mathrm{d}P/\mathrm{d}\lambda^k$，且$f$称为$P$的密度．

复数一般不常见，其定义见附录B. 如果f是一个复值函数，$f\equiv g + ih$，其中g和f是实值函数，那么对任意测度μ, $\int f\mathrm{d}\mu$定义为$\int g\mathrm{d}\mu + \mathrm{i}\int h\mathrm{d}\mu$.

设(x, t)是$\mathbb{R}^k$中向量x和t的通常内积，考虑任意取值于$\mathbb{R}^k$的随机变量X和法则P，X或P的特征函数由下式定义：

$$f_P(t) := Ee^{\mathrm{i}(X,t)} = \int e^{\mathrm{i}(x,t)}\mathrm{d}P(x), \forall t \in \mathbb{R}^k.$$

(回顾对任意实数u，$e^{\mathrm{i}u} = \cos(u) + \mathrm{i}\cdot\sin(u)$.)因此$f_P$是$P$的傅里叶变换，但是没有大多数傅里叶分析中用的常数因子$(2\pi)^{-k/2}$. 注意，对任意法则P，$f_P(0) = 1$. 例如，对$+1$和-1的法则P给定测度为1/2，则特征函数$f_P(t) \equiv \cos t$.

刚刚定义的“特征函数”解释了为什么概率统计工作者喜欢用1_A作为集合A的示性函数(而不像数学其余分支那样选用特征函数).

下面将给出$\mathbb{R}^1$中满足$EX_j < \infty$的独立同分布变量X_j的标准化部分和S_n的极限密度函数．

9.4.1 命题 *对任意$m\in\mathbb{R}$，$\sigma > 0$，函数*

$$x \mapsto \varphi(m,\sigma^2,x) := \frac{1}{\sigma\sqrt{2\pi}}\exp(-(x-m)^2/(2\sigma^2))$$

是$\mathbb{R}$上的概率密度函数．

证明 通过变量变换可假定$m = 0$，$\sigma = 1$(如果$u := (x-m)/\sigma$，那么$\mathrm{d}u = \mathrm{d}x/\sigma$)．运用Tonelli-

298 Fubini定理和极坐标(4.4节习题6)可得，

$$\left(\int_{-\infty}^{\infty}\exp(-x^2/2)\mathrm{d}x\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp((-x^2-y^2)/2)\mathrm{d}x\mathrm{d}y$$

$$= \int_0^{2\pi} d\theta \int_0^\infty r\exp(-r^2/2)\,dr$$

$$= 2\pi(-\exp(-r^2/2))\,|_0^\infty = 2\pi. \qquad \square$$

设 $N(m, \sigma^2)$ 是密度为 $\varphi(m, \sigma^2, \cdot)$ 的分布函数，其中 $m \in \mathbb{R}$，$\sigma > 0$. 在积分中，当 P 像在 $N(m, \sigma^2)$ 中那样依赖于其他参数时，$P(dx)$ 可用 $dP(x)$ 代替. 对 $N(m, \sigma^2)$ 的积分可以通过线性变量变换变成对 $N(0, 1)$ 的积分. 那么和上面对 r 的积分一样，有

$$\int_{-\infty}^{\infty} xN(0,1)(dx) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} -d(\exp(-x^2/2)) = 0.$$

然后运用分部积分和命题 9.4.1 可得，

$$\int_{-\infty}^{\infty} x^2N(0,1)(dx) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} -xd(\exp(-x^2/2)) = 1,$$

即得到了 $N(m, \sigma^2)$ 的均值和方差(换句话说，具有这种分布的任意随机变量的均值和方差)由下式给出：

$$\int_{-\infty}^{\infty} xN(m,\sigma^2)(dx) = m, \int_{-\infty}^{\infty} (x-m)^2N(m,\sigma^2)(dx) = \sigma^2.$$

分布函数 $N(m, \sigma^2)$ 称为 $\mathbb{R}$ 上的正态分布或高斯分布，其均值为 m、方差为 σ^2，$N(0, 1)$ 称为标准正态分布(standard normal distribution).

对 $\mathbb{R}$ 上的任意法则 P，积分 $\int_{-\infty}^{\infty} x^n dP(x)$（若有定义）称为 P 的矩(moment). 因此一阶矩就是均值，如果均值是0，二阶矩就是方差. 对所有 u 的值，函数 $g(u) = \int_{-\infty}^{\infty} e^{xu} dP(x)$ 是有定义且有限的，称为 P 的矩母函数(moment generating function). 如果它在0点邻域内有定义且有限，则 e^{xu} 在 $x=0$ 的泰勒展开给出了 P 的 n 阶矩它是 g 的 n 阶导数在0点的值，在下面的证明中将以 $P=N(0, 1)$ 为例给予说明. 从这个意义上讲，g 可以生成矩.

299

9.4.2 命题

(a) $\mathbb{R}$ 上正态分布的特征函数由下式给出：

$$\int_{-\infty}^{\infty} e^{ixu}N(m,\sigma^2)(dx) = \exp(imu - \sigma^2u^2/2).$$

(b) 对所有实数 u，

$$\int_{-\infty}^{\infty} e^{xu}N(0,1)(dx) = \exp(u^2/2).$$

(c)

$$\int_{-\infty}^{\infty} x^{2n}N(0,1)(dx) = (2n)!/(2^n n!), n = 0,1,\cdots,$$

$$= 1\cdot 3\cdot 5\cdots\cdot(2n-1), n \geq 1,$$

$$\int_{-\infty}^{\infty} x^m N(0,1)(dx) = 0, m = 1,3,5,\cdots.$$

证明 首先，对(b)，$xu - x^2/2 \equiv u^2/2 - (x-u)^2/2$，因此，(b)左边的积分等于 $\exp(u^2/2)N(0, 1)(\mathbb{R}) = \exp(u^2/2)$，证毕.

下面证(c)，由于 $\exp(|xu|) \leq \exp(xu) + \exp(-ux)$，$\exp(|xu|)$ 对 $N(0, 1)$ 是可积的. 对 $Z = |xu|$，e^z 的泰勒级数是正项级数，因此，由控制收敛定理知，它关于 $N(0, 1)(dx)$ 是逐项可积

的. 同理可知，e^{xu}(作为 x 的函数)的泰勒级数也是逐项可积的．函数 $g(u)$ 的幂级数展开式中 u^n 的系数是唯一的，为 $g^{(n)}(0)/n!$．展开(b)式的两边，在 e^z 的泰勒级数中令 $z=u^2/2$，且 u 的每次项的系数相等．奇次项系数是 0，所以 $N(0,1)$ 的奇阶矩是 0. 对偶次项，等式右边为$u^{2n}/(2^n n!)$，左边为 $N(0,1)$ 的第 $2n$ 阶矩的 $u^{2n}/(2n)!$ 倍，故(c)得证．

再证(a)，令 $v=(x-m)/\sigma$，因此 $e^{ixu}=e^{imu}e^{iv\sigma u}$. 这种替换可将一般情况归结为 $m=0$，$\sigma=1$. 我们仅需证明将(b)中的任意实数 u 用虚数 iu 代替结论仍成立即可．为此，根据 e^z 的泰勒级数展开 e^{ixu}，由控制收敛定理($|e^{ixu}|=1\leqslant e^{|xu|}$)可得此级数(实部或虚部)可以逐项积分，且对 $z=ixu$ 和 $z=|xu|$，e^z 的泰勒级数中相应项的绝对值是相等的．对每个 n，$(iu)^n$ 的系数由(c)给出，然后对级数求和便可证得(a). □

例：设 X 是一个服从二项分布的随机变量，即为 n 次独立试验成功的次数，其中每次成功的概率 p，因此，

300

$$P(X=k)=\binom{n}{k}p^k q^{n-k},k=0,1,\cdots,n,q:=1-p.$$

根据二项式定理可知，这个分布的矩母函数是 $\sum_{0\leqslant k\leqslant n}e^{kt}P(X=k)=(pe^t+q)^n$.

这个分布的独立实随机变量的和相应于它们法则的卷积(定理 9.1.3). 下面将给出它们特征函数积的相关简单运算．

9.4.3 定理 如果 X_1，X_2，…是$\mathbb{R}^k$ 中的独立随机变量，$S_n=X_1+X_2+\cdots+X_n$，则 $\forall t\in\mathbb{R}^k$，

$$E\exp(i(S_n,t))=\prod_{j=1}^n E\exp(i(X_j,t)).$$

因此，如果 X_j 是独立同分布的，且特征函数为 f，那么 S_n 的特征函数为 $f(t)^n$.

证明 由 $\exp(i(S_n,t))=\prod_{1\leqslant j\leqslant n}\exp(i(X_j,t))$、独立性和 Tonelli-Fubini 定理可得结论． □

有限绝对矩 $E|X|^r$ 的全局性质蕴涵了特征函数的局部微分性质．

9.4.4 定理 如果 f 是实随机变量，且对某个整数 $r\geqslant 0$，$E|X|^r<\infty$，那么特征函数 $f(t):=E\exp(iXt)$ 是 C^r 的，即 f 在$\mathbb{R}$ 上有 r 阶连续导数，其中对 $Q:=\mathcal{L}(X):=P\circ X^{-1}$，

$$f^{(j)}(t)=\int_{-\infty}^{\infty}(ix)^j e^{ixt}dQ(x),\quad j=0,1,\cdots,r,$$

$$f^{(j)}(0)=\int_{-\infty}^{\infty}(ix)^j dQ(x).$$

证明 当 $r=0$ 时，由于 e^{ixt}关于 t 是连续的且是一致有界的，因此由控制收敛定理可得 f 是连续的．当 $r=1$ 时，对任意 a 和 b，$g(u):=e^{iu}$，$|g(b)-g(a)|=\left|\int_a^b g'(t)dt\right|\leqslant|b-a|\sup_{a\leqslant x\leqslant b}|g'(x)|=|b-a|$，所以对任意的实数 x，t，h，$|e^{ix(t+h)}-e^{ixt}|\leqslant|xh|$. 由控制收敛定理知，在积分号下可对 f 微分．通过对第 r 阶导数迭代即可得出结论． □

9.4.5 定理 设 X_1，X_2，…是 i.i.d 实随机变量，满足 $EX_1=0$，$EX_1^2:=\sigma^2<\infty$．令 $S_n:=X_1+\cdots+X_n$，则对所有的实数 t，

301

$$\lim_{n\to\infty}E\exp(iS_n t/n^{1/2})=\exp(-\sigma^2t^2/2).$$

注：最后一个定理说明法则 $S_n/n^{1/2}$ 的特征函数在$\mathbb{R}$ 上逐点收敛到法则 $N(0,\sigma^2)$ 的特征函数．这是证明法则收敛的步骤之一．

证明 令 $f(t):=E\exp(iX_1t)$，那么由定理 9.4.4 知，$f(0)=1$，$f'(0)=0$，$f''(0)=\sigma^2$. 因此，由带余项的泰勒定理可知，当 $t\to 0$ 时，$f(t)=1-\sigma^2t^2/2+o(t^2)$，其中 o 的意思是当 $t\to 0$ 时，$o(t^2)/t^2\to 0$. 现在由定理 9.4.3 可得，对任意固定的 t，当 $n\to\infty$ 时，

$$E\exp(iS_nt/n^{1/2}) = f(t/n^{1/2})^n = (1-\sigma^2t^2/(2n)+o(t^2/n))^n \to \exp(-\sigma^2t^2/2).$$

当 $n\to\infty$ 时，取对数和它们的泰勒级数可得结论. □

习题

1. 计算：(a) $\int_{-\infty}^{\infty} x^2N(3,4)(dx)$；(b) $\int_{-\infty}^{\infty} x^4N(1,4)(dx)$.
2. 试求法则 P 的矩母函数和矩，其 P 的密度函数在 $x>0$ 时为 e^{-x}，在 $x<0$ 时为 0(指数分布).
3. 求习题 2 中指数分布的特征函数，并求对所有的 x 密度函数为 $e^{-|x|}/2$ 的律的特征函数. 对于后一种情况，像定理 9.4.5 中那样求 $S_n/n^{1/2}$ 的特征函数，并证明当 $n\to\infty$ 时是收敛的.
4. (a) 如果 X_j 是独立随机变量，其矩母函数为 $f_j(t)$，对某个常数 $h>0$，$|t|\leqslant h$，$f_j(t)$ 有定义且有限，证明：如果 $|t|\leqslant h$，S_n 的矩母函数是 $X_1,\cdots,X_n$ 的矩母函数的积.
   (b) 如果 X 是非负随机变量，g 为 X 的矩母函数，证明：对任意 $t\geqslant 0$，
   $$P(X\geqslant t)\leqslant \inf_{u\geqslant 0} e^{-tu}g(u).$$
   (c) 如同定理 9.4.3 前面的例子，$X(n,p)$ 是 n 次独立试验成功的次数，其中每次试验成功的概率为 p，运用结论(a)，求 $X(n,p)$ 的矩母函数. [提示：X 是 n 个简单独立变量的和.]
   (d) 证明：$E(k,n,p):=P(X(n,p)\geqslant k)\leqslant (np/k)^k(nq/(n-k))^{n-k}$，$k\geqslant np$. [提示：对 $t=k$，$e^u=kq/((n-k)p)$ 应用(b).]
5. 求下列法则的矩母函数.
   (a) P 服从泊松分布：对某个 $\lambda>0$，$P(\{k\})=e^{-\lambda}\lambda^k/k!$，$k=0,1,\cdots$.
   (b) P 服从 $[0,1]$ 上的均匀分布(勒贝格测度).

302

6. 试求 $\mathbb{R}$ 上关于勒贝格测度密度函数为 $f=1_{[-1,1]}/2$ 的法则 P 的特征函数.
7. 如同定理 9.4.3 前面的例子所述，设 X 服从二项分布，通过其矩母函数求 EX、EX^2 和 EX^3.
8. 设 f 是 $\mathbb{R}$ 上的复值函数，使得在 0 点的邻域内，$f'(t)$ 关于 t 是有限的且 $f''(0)$ 是有限的.
   (a) 证明：$f''(0)=\lim_{t\to 0}(f(t)-2f(0)+f(-t))/t^2$. [提示：利用带余项的泰勒级数.]
   (b) 如果 f 是法则 P 的特征函数，证明：$\int x^2dP(x)<\infty$. [提示：利用(a)和法图引理.]
9. 试举例：在 $\mathbb{R}$ 上特征函数为 f 的律 P，满足 $f'(0)$ 有限，但 $\int |x|dP(x)=+\infty$. [提示：设对适当的常数 c，P 有密度函数 $c1_{[3,\infty]}(|x|)/(x^2\log|x|)$. 利用 $1-\cos(x)t\leqslant x^2t^2/2(|x|\leqslant 2/|t|)$，$1-\cos(xt)\leqslant 2(|x|>2/|t|)$，证明 f 是实值函数且当 $|t|\to 0$ 时，$|1-f(t)|=o(|t|)$.]

9.5 特征函数的唯一性和中心极限定理

本节中 C_b 表示有界连续复值函数的集合. 首先给出了(如同定理 9.4.5 中所述的)特征函数收敛意味着分布收敛(定理 9.4.5)，并且分布函数和特征函数是 1－1 对应的.

9.5.1 唯一性定理 如果 P，Q 是 $\mathbb{R}^k$ 上的法则，且有相同的特征函数 g，那么 $P=Q$.

在习题 7～9 中，特征函数在 0 的邻域内是相等的但不是处处相等.

在证明定理 9.5.1 过程中，用律与正态分布作卷积是有帮助的. 设 $N(0,\sigma^2I)$ 是 $\mathbb{R}^k$ 上的法则，

它是 $N(0,\sigma^2)$ 的 k 次积，因此，坐标 $x_1,\cdots,x_k$ 是分布函数为 $N(0,\sigma^2)$ 的独立同分布随机变量．当 $\delta\downarrow 0$ 时，卷积 $P^{(\sigma)}:=P*N(0,\sigma^2 I)$ 将收敛到法则 P，这些卷积总是有密度函数，并且对正态分布可以明确地进行傅里叶分析（特征函数的计算）．具体的见下面两个引理．

9.5.2 引理 设 P 是 $\mathbb{R}^k$ 上的概率法则，其特征函数为 $g(t)=\int e^{i(x,t)}dP(x)$．那么 $P^{(\sigma)}$ 有密度函数 $f^{(\sigma)}$，且满足

303

$$f^{(\sigma)}(x)=(2\pi)^{-k}\int g(t)\exp(-i(x,t)-\sigma^2|t|^2/2)dt.$$

证明 设 φ_0 是 $N(0,\sigma^2 I)$ 的密度函数．首先假设 $k=1$，对 $\forall x\in\mathbb{R}$，可把 $f^{(\sigma)}$ 写为

$$\int\varphi_\sigma(x-y)dP(y) \qquad (由 9.1.6)$$

$$=\int\varphi_\sigma(y-x)dP(y) \qquad (\varphi_\sigma 是偶函数)$$

$$=\int(2\pi)^{-1}\int\exp(i(y-x)t-\sigma^2t^2/2)dtdP(y) \qquad (由 9.4.2a)$$

$$=(2\pi)^{-1}\int\int e^{iyt}dP(y)\exp(-ixt-\sigma^2t^2/2)dt$$

（对 $m=0$，将 σ 和 $1/\sigma$ 互换，由 Tonelli-Fubini 定理）

$$=(2\pi)^{-1}\int g(t)\exp(-ixt-\sigma^2t^2/2)dt.$$

对 $k>1$，其证明的本质是一样的，例如，用 (x,t) 来替换 xt，$|t|^2$ 替换 t^2，$(2\pi)^{-k}$ 替换 $(2\pi)^{-1}$． □

9.5.3 引理 对 $\mathbb{R}^k$ 上的任意法则 P，当 $\sigma\downarrow 0$ 时，$P^{(\sigma)}\to P$．

证明 设 X 有法则 P，在 $\mathbb{R}^k$ 上 Y 和 X 相互独立且有分布 $N(0,I)$．那么对每个 $\sigma>0$，σY 有分布 $N(0,\sigma^2 I)$，因此根据定理 9.1.3 知，$X+\sigma Y$ 有法则 $P^{(\sigma)}$，$X+\sigma Y$ 几乎必然收敛到 X，所以当 $\sigma\downarrow 0$ 时其依概率收敛．故由命题 9.3.5 可知，法则 $P^{(\sigma)}\to P$． □

唯一性定理 9.5.1 的证明 由引理 9.5.2 知，法则 $P^{(\sigma)}$ 有由 g 决定的密度函数，因此对所有的 $\sigma>0$，$P^{(\sigma)}=Q^{(\sigma)}$．再由引理 9.5.3 可得，当 $\sigma\downarrow 0$ 时，法则 $P^{(\sigma)}$ 分别收敛到 P 和 Q．由引理 9.3.2 收敛极限的唯一性可得 $P=Q$． □

由定理 9.5.1 知，法则对应特征函数的值是可逆的，对某些法则而言，其逆可由积分方程给出．

9.5.4 傅里叶反演定理 设 f 是 $\mathbb{R}^k$ 上的概率密度函数，g 是它的特征函数，且

304

$$g(t):=\int f(x)e^{i(x,t)}dx,\ \forall t\in\mathbb{R}^k.$$

如果 g 在 $\mathbb{R}^k$ 上关于勒贝格测度 dt 是可积的，那么对几乎所有的 $x\in\mathbb{R}^k$，

$$f(x)=(2\pi)^{-k}\int g(t)e^{-i(x,t)}dt.$$

例如，如同命题 9.4.2(a) 中表明的，正态密度函数 f 对 $k=1$ 有可积的特征函数，因此可使用定理 9.5.4. 另一方面，任意特征函数是连续的（根据控制收敛定理），同理，定理 9.5.4 最后一个积分表示对 g 可积的 x 的连续函数．因此，定理 9.5.4 仅能用于密度函数几乎处处连续的情况，对于多数密度函数，如均匀密度函数 $1_{[0,1]}$ 和指数密度函数 $e^{-x}1_{[0,\infty)}(x)$，这个定理都不成立．

证明 在引理 9.5.2 中，当 $\sigma\downarrow 0$ 时，等式右边关于 x 一致收敛到 $h(x):=(2\pi)^{-k}\int g(t)e^{-i(x,t)}dt$，由

于 g 是可积的，且由控制收敛定理，

$$\int |g(t)||1-\exp(-\sigma^2|t|^2/2)|dt \to 0.$$

对$\mathbb{R}^k$上具有紧支集的任意连续函数 v，当 $\sigma\downarrow 0$ 时，$\int v dP^{(\sigma)}$ 收敛到 $\int vh dx$，由引理9.5.3知，它也收敛到 $\int vf dx$，因此这两个积分相等．正如引理9.3.2中那样，对每一有界开集 U，或任意有界闭集，或任意有界博雷尔集，$\int_U f dx = \int_{Uh} dx$. 对每个 m，限制于 $|x|<m$，从而 f 和 h 关于勒贝格测度是相同测度两个密度函数，因此由 Radon-Nikodym 定理(5.5.4)的唯一性知，对于 $|x|<m$ 几乎处处有 $f=h$. 即在$\mathbb{R}^k$上几乎处处有 $f=h$. □

9.8 节将证明特征函数收敛于连续极限函数表明了它是一致紧的．下面给出一个充分条件．

9.5.5 引理 设 P_n 是$\mathbb{R}^k$上一致紧的法则序列，对所有的 t，其特征函数 $f_n(t)$ 收敛到 $f(t)$，那么 $P_n\to P$，其中 P 的特征函数为 f.

证明 由定理9.3.3知，P_n 的任意子序列都有一个收敛到某法则的子子序列．所有法则的极限都有特征函数 f，由唯一性定理知，所有法则极限的特征函数都与某法则 P 的相等．那么由命题9.3.1可知，$P_n\to P$. □

305

例：法则 $N(0, n)$ 的特征函数 $f_n := \exp(-nt^2/2)$ 不是一致紧的，也不是收敛的，但对所有的 t，$f_n(t)$ 收敛到 $1_{\{0\}}(t)$. 这表明在引理中一致紧性是必需的．

对取值于$\mathbb{R}^k$的随机变量 $Y=\langle Y_1, \cdots, Y_k\rangle$ 和法则 P，Y 或 P 的均值(或期望)定义为 $EY:=\langle EY_1, \cdots, EY_k\rangle$ 当且仅当这些均值存在且是有限的．因此有 $|Y|^2 := Y_1^2+\cdots+Y_k^2$. 一般地，令 $S_n := X_1+\cdots+X_n$. 回忆(8.1节)对任意两个实随机变量 X 和 Y，且 $EX^2<\infty$，$EY^2<\infty$，则 X 和 Y 的协方差由下式定义：

$$\operatorname{cov}(X,Y) := E((X-EX)(Y-EY)) = E(XY) - EXEY.$$

因此，X 与其自身的协方差是它的方差．如果 X 是取值于$\mathbb{R}^k$的随机变量且 $E|x|^2<\infty$，那么 X 或它的法则的协方差阵(covariance matrix)是 $k\times k$ 矩阵 $\operatorname{cov}(X_i, X_j)$，其中 $i, j=1, \cdots, k$. 对任意常数向量 V，很显然 X 和 $X+V$ 有相同的协方差．

例：设 P 是$\mathbb{R}^2$上的法则且 $P((0, 0))=P((0, 1))=P((1, 1))=1/3$，那么 P 的均值为(1/3, 2/3)，协方差阵为

$$\begin{bmatrix} 2/9 & 1/9 \\ 1/9 & 2/9 \end{bmatrix}.$$

下面介绍概率论中的一个主要定理，更经典的一维情况将在定理的第二部分给出．

9.5.6 中心极限定理

(a) 设 $X_1, X_2, \cdots$ 是取值于$\mathbb{R}^k$的独立同分布随机变量，$EX_1=0$，$0<E|X_1|^2<\infty$. 那么当 $n\to\infty$ 时，$\mathcal{L}(S_n/n^{1/2})\to P$，其中 P 的特征函数为

$$f_P(t) = \exp\left(-\frac{1}{2}\sum_{r,s=1}^{k} C_{rs}t_r t_s\right), C_{rs} := EX_{1r}X_{1s}$$

(C 是随机向量 X_1 的协方差矩阵).

306

(b) 如果(a)的假设成立，$k=1$，$EX^2=1$，那么对任意实数 x，当 $n\to\infty$ 时，

$$P(S_n/n^{1/2} \leqslant x) \to \Phi(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} \exp(-t^2/2)\,dt.$$

证明 向量 X_i 是独立的，且均值为 0，因此对 $i \neq j$，$E(X_i, X_j) = 0$. 对所有的 n，对 $k=1$ 运用定理 8.1.2，得 $E|S_n/n^{1/2}|^2 = E|X_1|^2$. 给定 $\varepsilon > 0$，对足够大的 M，$E|X_1|^2/M^2 < \varepsilon$. 根据切比雪夫不等式(8.3.1)知，对所有的 n，$P(|S_n/n^{1/2}| > M) < \varepsilon$，故 $S_n/n^{1/2}$ 的是一致紧的.

现在对每一 $t \in \mathbb{R}^k$，(t, X_i) 是独立同分布的实值随机变量，且均值为 0，$E(t, X_1)^2 < \infty$. 因而由定理 9.4.5 知，对所有的 $t \in \mathbb{R}^k$，

$$E\exp(i(t, S_n/n^{1/2})) \to \exp(-E(t, X_1)^2/2), n \to \infty,$$

其中右边的函数等于 f_p. 因此由引理 9.5.5 知，对特征函数为 f_p 的法则 P，$\mathcal{L}(S_n/n^{1/2}) \to P$，故(a)得证. 对于(b)，可用定理 9.3.6 的方法，因为对所有的 x，Φ 是连续的，然后根据唯一性定理 9.5.1 和命题 9.4.2(a)($N(0, 1)$ 的特征函数是 $\exp(-t^2/2)$)即可得证. □

对任意 r, s，协方差矩阵 C 是对称的：$C_{rs} = C_{sr}$. C 是非负定的，即 $\forall t \in \mathbb{R}^k$，

$$\sum_{r,s=1}^{k} C_{rs}t_r t_s = E\left(\sum_{r=1}^{k} t_r X_{1r}\right)^2 \geqslant 0.$$

例：设

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 5 & -2 \\ -2 & 4 \end{bmatrix}.$$

那么 A 和 B 不是协方差矩阵，因为 B 不是对称的，A 不是非负定的：对 $t = (-1, 1)$，$(At, t) < 0$. C 是协方差阵.

$\mathbb{R}^k$ 上具有由中心极限定理给出的特征函数的概率法则称为 $N(0, C)$，即均值为 0，方差为 C 的正态分布. 下面定理证明了这样的分布确实存在且方差为 C.

307

9.5.7 定理 *对任意 $k \times k$ 的非负定对称矩阵 C，$\mathbb{R}^k$ 上的概率法则 $N(0, C)$ 存在，均值为 0，协方差为*

$$\int x_r x_s \, dN(0, C)(x) = C_{rs}, \qquad r, s = 1, \cdots, k.$$

设

$$H := \text{range}C = \left\{ \left\{ \sum_{s=1}^{k} C_{rs}y_s \right\}_{r=1}^{k} : y \in \mathbb{R}^k \right\}.$$

那么 $N(0, C)(H) = 1$. 设 j 是 H 的维数，因此 $j = \text{rank}(C)$. 由 C 定义的线性变换 D 限定在 H 上，给出 H 映上到自身的线性变换 D_H. 对可对角化矩阵 C，$N(0, C)$ 关于勒贝格测度的密度函数(Radon-Nikodym 导数)由下式给出：

9.5.8
$$(2\pi)^{-j/2}(\det D_H)^{-1/2}\exp(-(D_H^{-1}x, x)/2), \qquad x \in H.$$

例：对 $k=1$，回顾(命题 9.4.1 和命题 9.4.2)密度函数为 $(2\pi)^{-1/2}\sigma^{-1}\exp(-x^2/(2\sigma^2))$ 的分布 $N(0, \sigma^2)$ 的特征函数为 $\exp(-\sigma^2 t^2/2)$. 这些法则的乘积即为(9.5.8)的情形，其中 D_H 是对角矩阵.

证明 因为 C 是实对称矩阵，D 是自伴的，即对所有 $x, y \in \mathbb{R}^k$，$(Dx, y) = (x, Dy)$. 如果 $z \perp H$，即对所有 $h \in H$，$(z, h) = 0$，那么对所有的 x，$0 = (Dx, z) = (x, Dz)$，因此 $Dz = 0$(取 $x = Dz$). 反之，如果 $Dz = 0$，那么 $z \perp H$，因此 $H^{\perp} := \{z: z \perp H\} = \{z: Dz = 0\}$.

现在 D 是自伴的，因此有特征向量 e_i(即 $D(e_i) = d_i e_i$，$i = 1, \cdots, k$)的一组规范正交基(参考

Hoffman 和 Kunze，1971，p314，定理18）. 满足 $d_i \neq 0$，因而满足 $d_i > 0$ 的特征向量 e_i 恰是 H 中的特征向量，并构成 H 的一组规范正交基．设其为 e_1，…，e_j. 将 D 限定在 H 上，可得 H 映射到自身的变换 D_H，因为 $d_j > 0(i=1, \cdots, j)$，故 D_H 是 H 映上到自身的变换．对给定的基，D_H 可表示对角线上是 d_i 的对角矩阵．

考虑一维的情况，由命题9.4.1知，对任意 $\sigma > 0$，在$\mathbb{R}$上存在一分布 $N(0, \sigma^2)$，其关于勒贝格测度的密度函数为 $(2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$．由命题9.4.2知，它的特征函数为 $\exp(-\sigma^2 t^2/2)$. 对 $\sigma=0$，$N(0, 0)$定义为在0处的点质量 δ_0，它的特征函数是常数1. 现在对于基 $\{e_r\}_{1\leq r\leq j}$，可用$\mathbb{R}^j$表示 H. 设 P 是法则 $N(0, d_r)$的笛卡儿积，其中 $r=1, \cdots, k$. 因为对 $r>j$，$d_r=0$，则由 Tonelli-Fubini 定理得，P 的特征函数为 308

$$f_P(t) = \exp\left(-\sum_{r=1}^{j} d_r t_r^2/2\right) = \exp(-(Dt,t)/2).$$

因而 P 有分布 $N(0, C)$的特征函数，因此 $N(0, C)$是存在的并根据唯一性定理可得 $P=N(0, C)$.

在 H 上，对于给定的基$\{e_r\}_{1\leq r\leq j}$，D_H^{-1} 可表示为对角线上第 r 个元素是 $1/d_r(r=1, \cdots, j)$的对角矩阵，D_H 的行列式是 $\prod_{1\leq r\leq j} d_j$. 因而由定义和命题9.4.1知，$P$ 有由式(9.5.8)给出的密度函数．

为了求得 P 的均值和协方差，我们可再次运用 D 对角化的基，以及前面命题9.4.2处理一维情形的方法．显然均值为0，且在对坐标对角化过程中，

$$\int (x,y)(x,z)\,\mathrm{d}N(0,C)(x) = (Dy,z)$$

因此对其他的基，有着与原有结果类似的结论，因而定理9.5.7得证． □

对$\mathbb{R}^k$上任意法则 P 和 $t\in\mathbb{R}^k$，P_t 是将 P 平移 t，这里对任意可测集 A，$P_t(A) := P(A-t)$，其中 $A-t := \{a-t: a\in A\}$. 如果随机变量 X 的法则为 P，那么 $X+t$ 的法则为 P_t. 因此，如果 $EX=m$，则 $\int y\,\mathrm{d}P_t(y) = E(X+t) = m+t$. 注意 $P_t = P * \delta_t$，那么对特征函数有

$$\int \mathrm{e}^{\mathrm{i}(x,y)}\mathrm{d}P_t(x) = \int \mathrm{e}^{\mathrm{i}(x,y)}\mathrm{d}P(x-t) = \int \mathrm{e}^{\mathrm{i}(u+t,y)}\mathrm{d}P(u)$$
$$= \mathrm{e}^{\mathrm{i}(t,y)}\int \mathrm{e}^{\mathrm{i}(u,y)}\mathrm{d}P(u).$$

因此，正如一维正态分布一样，变换是由 $\mathrm{e}^{\mathrm{i}(t,y)}$ 的特征函数(记作 y 的函数)与 t 的乘积得到．

对所有的 x，记 $P_t = P\circ\tau_t^{-1}$，其中 $\tau_t(x) := t+x$. 对任意的 $m\in\mathbb{R}^k$ 和法则 $N(0, C)$，$N(m, C)$是由 $N(0, C)$平移 m 而得到的．换句话说，$N(m, C) := N(0, C)\circ\tau_m^{-1}$. 因此，如果随机变量 X 有法则 $N(0, C)$，那么 $X+m$ 有法则 $N(m, C)$. 这里 $N(m, C)$的均值为 m，协方差为 C，称为*均值为 m、协方差为 C 的正态分布*(normal law with mean m and covariance C).

笛卡儿积和法则的卷积的连续性不用特征函数也可以证明，但利用特征函数证明更为方便． 309

9.5.9 定理 *对任意正整数 m 和 k，如果 P_n 是$\mathbb{R}^m$上的法则，Q_n 是$\mathbb{R}^k$上的法则，且 $P_n\to P_0$，$Q_n\to Q_0$. 那么笛卡儿积法则收敛，即在$\mathbb{R}^{m+k}$上，$P_n\times Q_n\to P_0\times Q_0$. 如果 $k=m$，那么在$\mathbb{R}^k$上，$P_n * Q_n\to P_0 * Q_0$.*

证明 首先，根据命题9.3.4知，P_n 在$\mathbb{R}^m$上是一致紧的，同理，Q_n 在$\mathbb{R}^k$上也是一致紧的．给定 $\varepsilon>0$，取紧集 J 和 K，使得对所有的 n，$P_n(J) > 1-\varepsilon/2$，$Q_n(K) > 1-\varepsilon/2$. 那么 $J\times K$ 是紧的，且对所有的 n，$(P_n\times Q_n)(J\times K) > 1-\varepsilon$，因此，法则 $P_n\times Q_n$ 是一致紧的．$\mathbb{R}^{k+m}$上的每个函数 $\mathrm{e}^{\mathrm{i}(t,x)}$

是$\mathbb{R}^k$和$\mathbb{R}^m$上此类函数的乘积．因而由 Tonelli-Fubini 定理知，$P_n \times Q_n$ 上的特征函数是 P_n 上的特征函数和 Q_n 上的特征函数的乘积，并且收敛到 $P_0 \times Q_0$. 那么由引理 9.5.5 得，$P_n \times Q_n \to P_0 \times Q_0$.

下面证明卷积的连续性，$P_n * Q_n$ 是$\mathbb{R}^{2k}$上 $P_n \times Q_n$ 的象法则，且$\langle x, y\rangle \mapsto x+y$，因此，根据连续映射定理(9.3.7)即可完成证明．□

对于与任意其他法则 P 卷积，δ_0 充当了单位元的角色，在这个意义上有 $\delta_0 * P = P * \delta_0 = P$，对其他的任意法则 μ 除了 δ_0 外没有别的法则可作为单位元，正如下面所示(尽管一般对卷积没有唯一的因式分解).

9.5.10 命题 *如果 μ 和 Q 是$\mathbb{R}$上的法则，那么 $\mu * Q = \mu$ 当且仅当 $Q = \delta_0$.*

证明 充分性是显然的，现证必要性：对所有的 t，特征函数 $f_\mu(t) f_Q(t) = f_\mu(t)$. 因为 f_μ 是连续的(根据定理 9.4.4，取 $r=0$)和 $f_\mu(0)=1$，所以存在 0 的一个邻域 U，在这个邻域内，$f_\mu(t) \neq 0$，因此 $f_Q(t)=1$.

$f_Q(t) = 1 (t \neq 0)$ 蕴涵着 $\int \cos(tx) \mathrm{d}Q(x) = 1$，并且由于对所有的 x，$\cos(tx) \leqslant 1$，故 $\cos(tx)=1$，因此对所有的 x，$x/2\pi$ 是一个整数，$Q(2\pi\mathbb{Z}/t) = 1$. 对邻域 U 中任意其他的 u，同样有 $Q(2\pi\mathbb{Z}/u) = 1$. 用 u/t 代替 u，得到 $Q = \delta_0$.□

下面将证明$\mathbb{R}^k$上的勒贝格测度 λ^k 在适当的线性变换下是不变的．从$\mathbb{R}^k$映射到自身的线性变换 T 称为*正交的*(orthogonal)，当且仅当对$\mathbb{R}^k$中所有的 x，y，$(Tx, Ty) = (x, y)$，其中$(\cdot,\cdot)$是$\mathbb{R}^k$中通常的内积．例如，在$\mathbb{R}^2$中，正交变换可能是旋转(围绕原点)或者是反射(在通过原点的直线上). 众所周知，按照几何学派的说法，通常的面积测度 λ^2 在正交变换下是不变的，但是实际证明可能不是平凡的．这里对 k 维的情形利用概率有一个相对简单的证明．

310

9.5.11 定理 *对任意从$\mathbb{R}^k$映射到自身的正交变换 T，$\lambda^k \circ T^{-1} = \lambda^k$，其中 $k=1, 2, \cdots$.*

证明 对 $k=2$，4.4 节的习题 6(极坐标)给出了一个证明，并且在命题 9.4.1 的证明中已经用于正态密度函数．

线性变换 T 的转置(伴随)为 T'，满足对所有的 x 和 y，$(Tx, y) = (x, T'y)$. T 为正交变换当且仅当 $T'T = I$(单位元). 因为 k 是有限的，这表明 T 映上的并且 $T' = T^{-1}$，即 $T = (T')^{-1}$，因此 $T'T = I$，且 T'也是正交的．由像测度定理(4.1.11)知，对 $N(0, I) \circ T^{-1}$ 的特征函数，

$$\begin{aligned}\int \mathrm{e}^{\mathrm{i}(u,y)} \mathrm{d}N(0,I) \circ T^{-1}(y) &= \int \mathrm{e}^{\mathrm{i}(u,Tx)} \mathrm{d}N(0,I)(x) \\ &= \int \mathrm{e}^{\mathrm{i}(T'u,x)} \mathrm{d}N(0,I)(x) = \exp(-|T'u|^2/2) = \exp(-|u|^2/2).\end{aligned}$$

由特征函数的唯一性(9.5.1)知，$N(0, I) \circ T^{-1} = N(0, I)$ ($N(0, I)$是正交不变的)，现对任意博雷尔集 B，且 $T^{-1}(B) = A$，应用定理 4.1.11 得，

$$\begin{aligned}\lambda^k \circ T^{-1}(B) &= (2\pi)^{k/2} \int_A \exp(|x|^2/2) \mathrm{d}N(0,I)(x) \\ &= (2\pi)^{k/2} \int 1_B(Tx) \exp(|Tx|^2/2) \mathrm{d}N(0,I)(x) \\ &= (2\pi)^{k/2} \int 1_B(y) \exp(|y|^2/2) \mathrm{d}N(0,I)(y) \\ &= \lambda^k(B).\end{aligned}$$

□

下面得出通过仿射变换，正态法则的像是正态的．

9.5.12 命题 设 $N(m, C)$ 是 $\mathbb{R}^k$ 上的正态法则，A 是 $\mathbb{R}^k$ 映射到某 $\mathbb{R}^j$ 的仿射变换，对从 $\mathbb{R}^k$ 映射到 $\mathbb{R}^j$ 的某线性变换 L 和 $w \in \mathbb{R}$，$x \in \mathbb{R}^k$，有 $Ax = Lx + w$. 那么 $N(m, C) \circ A^{-1}$ 是 $\mathbb{R}^j$ 的正态法则，特别地 $N(u, LCL')$，其中 $u := Lm + w$.

注：这里 δ_m 当作 $N(m, 0)$ 是可能的，例如，如果 $C = 0$ 或 $L = 0$. L 可由 $j \times k$ 矩阵矩阵表示，因此它的转置 L' 是一个 $k \times j$ 矩阵，对 $s = 1, \cdots, j$ 和 $r = 1, \cdots, k$，$(L')_{rs} = L_{sr}$.

311

证明 设 y 是 $\mathbb{R}^k$ 中的随机变量，其法则为 $N(0, C)$ 且 $x = y + m$. $t \in \mathbb{R}^j$ 时考虑其特征函数，

$$E\exp(\mathrm{i}(t, Lx + w)) = \exp(\mathrm{i}(t, u))E\exp(\mathrm{i}(L't, y)).$$

由定理 9.5.6 和定理 9.5.7 知，y 有特征函数

$$E\exp(\mathrm{i}(v, y)) = \exp(-v'Cv/2).$$

对 $v = L't$，$v'Cv = (L't)'CL't = t'LCL't$，因此（由特征函数的唯一性）$N(m, C) \circ A^{-1} = N(u, LCL')$，其中 LCL' 是对称非负定的 $j \times j$ 矩阵． □

如果 $(X_1, \cdots, X_k)$ 在 $\mathbb{R}^k$ 上有正态法则 $N(m, C)$，那么 $X_1, \cdots, X_k$ 称为有正态联合分布或是联合正态的．

9.5.13 定理 $\mathbb{R}^k$ 中的随机变量 $X = (X_1, \cdots, X_k)$ 有正态分布，当且仅当对每一 $t \in \mathbb{R}^k$，$(t, X) = t_1X_1 + \cdots + t_kX_K$ 在 $\mathbb{R}$ 中有正态分布．

证明 必要性在命题 9.5.12 中令 $j = 1$ 即可得．下证充分性：首先注意到，取 t 为 $\mathbb{R}^k$ 的通常基向量时，$E|x|^2 = EX_1^2 + \cdots + EX_k^2 < \infty$. 因而 X 有明确定义，即有限的协方差阵 C. 假设 $EX = 0$，由于 (X, t) 在 $\mathbb{R}$ 上有正态分布，其均值为 0，方差为 (Ct, t)，故特征函数为 $E\exp(\mathrm{i}(t, X)) = \exp(-(Ct, t)/2)$. 根据特征函数的唯一性（定理 9.5.1）和正态特征函数的形式（定理 9.5.6 和定理 9.5.7）可得 X 有正态分布． □

对联合正态变量，独立性和协方差为 0 是等价的．

9.5.14 定理 设 $(X, Y) = (X_1, \cdots, X_k, Y_1, \cdots Y_m)$ 在 $\mathbb{R}^{k+m}$ 中是联合正态的，那么 $X = (X_1, \cdots, X_k)$ 和 $(Y_1, \cdots, Y_m)$ 是独立的，当且仅当对所有的 $i = 1, \cdots, k$ 和 $j = 1, \cdots, m$，$\mathrm{cov}(X_i, X_j) = 0$，因此，联合正态实随机变量 X 和 Y 是独立的，当且仅当 $\mathrm{cov}(X, Y) = 0$.

证明 如定理 8.1.2 前面所提到的，平方可积的独立变量协方差为 0. 反之，对所有的 i 和 j，如果 (X, Y) 有联合正态分布 $N(m, C)$ 且 $\mathrm{cov}(X_i, X_j) = 0$，设 $m = (\mu, \nu)$，$EX = \mu$，$EY = \nu$，考虑特征函数 $f(t, u) := E\exp(\mathrm{i}((t, u) \cdot (X, Y)))$. 正如定理 9.5.7 证明后的注所示，我们有 $f(t, u) \equiv \exp(\mathrm{i}(t\mu + u\nu))g(t, u)$，其中 $g(t, u)$ 是 $N(0, C)$ 的特征函数．由于协方差为 0，所以 C 的形式为

312

$$C = \begin{bmatrix} D & 0 \\ 0 & F \end{bmatrix}$$

其中 D 是 X 的 $k \times k$ 协方差矩阵，F 是 Y 的 $m \times m$ 协方差矩阵．由正态特征函数的形式（定理 9.5.6 和定理 9.5.7）知，$g(t, u) \equiv h(t)j(u)$，其中 h 是 $N(0, D)$ 的特征函数，j 是 $N(0, F)$ 的特征函数．同理 $f(t, u) \equiv r(t)s(u)$，其中 r 是 $N(\mu, D)$ 的特征函数，s 是 $N(\nu, F)$ 的特征函数．根据特征函数的唯一性（定理 9.5.1）可得，$N(m, C) = N(\mu, D) \times N(\nu, F)$，也就是说，$X$ 和 Y 是独立的． □

习题

1. 设 X 和 Y 是独立实随机变量且 $\mathcal{L}(X) = N(m, \sigma^2)$，$\mathcal{L}(Y) = N(\mu, \tau^2)$，证明：$\mathcal{L}(X+Y) = N(m+\mu, \sigma^2+\tau^2)$.

2. ($\mathbb{Z}^2$ 中的随机游动). 设 X_1, X_2, …是取值于$\mathbb{Z}^2$ 的独立同分布的随机变量, 即 $X_j=(k_j, m_j)$, $k_j\in\mathbb{Z}$, $m_j\in\mathbb{Z}$. 那么记 S_n 为从(0, 0)点开始 n 步后随机游动的位置, 其中 X_j 是第 j 步游动. 试求在下列情况下, 当 $n\to\infty$ 时, 法则 $S_n/n^{1/2}$的极限, 并给出极限法则的密度函数.

   (a) X_1 的取值有四种可能: (1, 0), (0, 1), (−1, 0), (0, −1), 且每个取值的概率为1/4.

   (b) X_1 有9种可能的取值, 每个取值的概率是1/9, (k, m)中每个 k, m 取值为 −1, 0或1.

   (c) X_1 有法则 P 且 $P((-1,0))=P((1,0))=1/3$, $P((-1,1))=P((1,-1))=1/6$, 试求在此情况下协方差阵的特征向量.

3. 设 X_1, X_2, …, 是$\mathbb{R}^k$ 上的独立同分布随机变量, 且 $EX_1=0$, $0<E|X_1|^2<\infty$. 设 u_j 是独立同分布随机变量且和所有的 X_i 是独立的, 满足 $P(u_1=1)=p=1-P(u_1=0)$, $0<p<1$. 对所有的 j, 设 $Y_j=u_jX_j$, $T_n:=Y_1+\cdots+Y_n$, 当 $n\to\infty$ 时, 法则 $S_n/n^{1/2}$和 $T_n/n^{1/2}$的极限有何关系?

4. 设 X_1, X_2, …是独立同分布的且是区间[−1, 1]上的均匀分布, 其密度函数为 $1_{[-1,1]}/2$,

   (a) 设 $S_n=X_1+\cdots+X_n$, 试求 S_2 和 S_3 的密度函数(显然, 不仅是卷积积分).

313 (b) 当 $n\to\infty$ 时, 求法则 $S_n/n^{1/2}$的极限.

5. 设一个试验有三种不相容的结果 A, B, C, 分别有非零概率 p, r, s, 且 $p+r+s=1$. 在 n 次独立重复的试验中, 设 n_A 为 A 发生的次数. 证明: 当 $n\to\infty$ 时, $\mathcal{L}((n_A-np, n_B-nr, n_C-ns)/n^{1/2})$收敛. 试求在平面 $x+y+z=0$ 上关于勒贝格测度 $\mu=\lambda^2\circ T^{-1}$的极限法则的密度函数, 其中 $T(x, y)=(x, y, -x-y)$.

6. 设 P 为一法则且对有限集 F, $P(F)=1$. 对某法则 μ, ρ, 假设 $P=\mu*\rho$. 证明: 存在有限集 G 和 H, 使得 $\mu(G)=\rho(H)=1$.

7. (a) 求$\mathbb{R}$上密度函数为 $\max(1-|x|, 0)$的法则的特征函数.

   (b) 证明: $g(t):=\max(1-|x|, 0)$是$\mathbb{R}$上某法则的密度函数. [提示: 利用(a)和定理9.5.4.]

8. 设 h 是$\mathbb{R}$上周期为 2π 的连续函数, 且对所有的 t, $h(t)=h(t+2\pi)$, $h(0)=1$. 证明: h 是$\mathbb{R}$上法则 P 的特征函数, 当且仅当所有的傅里叶系数(Fourier coefficient) $a_n=\frac{1}{2\pi}\int_0^{2\pi}h(t)e^{int}dt$ $(n=\cdots -1, 0, 1, \cdots)$是非负的, 然后证明 $P(\mathbb{Z})=1$(假定5.4节习题6的结论成立).

9. 设 g 是习题7(b)中的函数(帐篷函数), 设 $h\mathbb{R}$ 上周期为2的周期函数且对 $|t|\leqslant 1$, 满足 $h(t)=g(t)$(锯齿函数). 证明: h 是特征函数. [提示: 应用习题8作从 2π 到2的尺度变换.] 注: y 和 h 是两个不同的特征函数, 但在[−1, 1]上是相等的.

10. 法则 δ_x 称为点群(point masses). 若概率法则 P 能唯一地表示成两个法则的卷积 $P=\mu*\rho$, 则称 P 是不可分解的, 其中 μ 和 ρ 是点式群体. 设 $P=p\delta_u+r\delta_v+s\delta_w$, $u<v<w$. 求在 p, r, s, u, v, w 满足什么条件下, P 是不可分解的? [提示: 如果 P 是不可分解的, 证明 $P=\mu*\rho$, 其中 μ 和 ρ 每个都集中在两个点上, 且 $v=(u+w)/2$. 那么 p, r, s 满足的条件是, 它们的二次多项式是有非负系数的线性多项式的乘积所应满足的条件.]

11. 设 f 是$\mathbb{R}$上法则 P 的特征函数且 $|f(s)|=|f(t)|=1$, 其中 $s/t(t\neq 0)$是无理数. 证明: 对某个 c, $P=\delta_c$. [提示: 证明对 P-几乎所有的 x, $e^{isx}=f(s)$. 参考命题9.5.10的证明.]

314 12. 设 H 为 k 维实希尔伯特空间且有一组正交基 e_1, …, e_k, 因此对每个 $x\in H$, $x=x_1e_1+\cdots+x_ke_k$. 在 H 上用 $d\lambda_H=dx_1\cdots dx_k$ 定义勒贝格测度 λ_H, 证明: λ_H 不依赖于基的选择(定理9.5.11的一个推论). 设 J 是另一个 k 维实希尔伯特空间, T 是从 H 映上到 J 的一个线性变换. 定义 $|\det T|$ 是 H 和 J 的某些正交基构成的矩阵 T 的行列式的绝对值. 证明: $\lambda_H\circ T^{-1}=|\det T|\lambda_J$, 使得 $|\det T|$ 不依赖于基的选择. [提示: 定义 T'从 J 映射到 H, 使得对所有的$x\in H$, $y\in J$, $(Tx, y)_J=(x, T'y)_H$. 得到 H 的组基使得 $T'T$ 是可对角化的. 回忆线性代数, 对任意 $k\times k$ 矩阵 A, B, $\det(AB)=\det(A)\det(B)$.]

13. 对$\mathbb{R}$上任意可测复值函数 f, 满足 $\int_{-\infty}^{\infty}|f|+|f|^2dx$ 是有限的. 令 $(Tf)(y)=(2\pi)^{-1/2}\int_{-\infty}^{\infty}f(x)e^{ixy}dx$. 证明: T 的定

义域和值域是复空间 $L^2(\mathbb{R},\lambda)$ 的稠密子集，对任意 T 的定义域中的 f，g，$\int(Tf)(y)(Tg)(y)^*\,dy=\int f(x)g(x)^*\,dx$，其中 * 表示复共轭．[提示：证明正态密度函数的线性组合满足这个等式，然后用其来逼近其他函数(Plancherel 定理)．]

14. 证明：X_1，X_2 为正态分布，但不是联合正态的．[提示：令对 $x_1=x_2$，$x_1=-x_2$，联合分布的测度都为 1/2.]
15. 如果 $X_1,\cdots,X_k$ 是相互独立的且服从分布 $N(0,1)$，所以 $X=(X_1,\cdots,X_k)$ 在 $\mathbb{R}^k$ 上的分布为 $N(0,I)$，那么 $\chi_k^2:=|X|^2=X_1+\cdots+X_k$ 为自由度为 k 的卡方分布．证明：$\chi_2^2/2$ 服从标准指数分布($t>0$ 时密度函数为 e^{-t}；$t<0$ 时密度函数为 0)．[提示：利用极坐标．]

9.6 三角形阵列和林德伯格定理

通常假设测量误差服从正态分布，至少近似服从正态分布．这个误差近似于独立同分布随机变量的和并不显然．在很多情况下，误差近似于独立但不一定同分布的随机变量的和．本节将证明满足某些条件的独立但不是同分布的变量的中心极限定理．

一个三角形阵列(triangular array)是随机变量 $X_{nj}(n=1,2,\cdots,j=1,\cdots,k(n))$ 的一个指标集，对给定的 n 值，阵列的行由 X_{nj} 组成．行和(row sum)这里定义为 $S_n:=\sum_{1\le j\le k(n)}X_{nj}$. 我们的目的是在适当的条件下证明当 $n\to\infty$ 时，法则 $\mathcal{L}(S_n)$ 是收敛的．这里假定每行内的随机变量是实值的且相互独立．因为所要讨论的仅仅是法则的收敛性，所以不同行之间的随机变量不必相互独立，甚至可以定义在不同的概率空间上．在 9.5 节中的正态变量的部分和可以由三角形阵列来定义：$k(n)=n$，$X_{nj}:=X_j/n^{1/2}$，$j=1,\cdots,n$，$n=1,2\cdots$.

315

如果每行仅包含一个非零项($k(n)=1$ 或对某个 i，$j\ne i$，$X_{nj}=0$)，那么 $\mathcal{L}(S_n)$ 可以是任意的．为了证明 $\mathcal{L}(S_n)$ 的收敛性，假定对足够大的 n，每项 X_{nj} 对第 n 行的和 S_n 影响很小．因此，每一行的长度 $k(n)$ 必趋于∞．这就是阵列为三角形(特别 $k=n$ 时)所给予的信息．下面是中心极限定理在三角形阵列的推广．

9.6.1 林德伯格定理 *假设对每个固定的 $n=1,2,\cdots$，$X_{nj}(j=1,\cdots,k(n))$ 是独立实随机变量，$EX_{nj}=0$，$\sigma_{nj}^2:=EX_{nj}^2$ 且 $\sum_{1\le j\le k(n)}\sigma_{nj}^2=1$. 令 $\mu_{nj}:=\mathcal{L}(X_{nj})$. 对任意 $\varepsilon>0$，令*

$$E_{nj\varepsilon}:=\int_{|x|>\varepsilon}x^2\,d\mu_{nj}(x).$$

假设对每个 $\varepsilon>0$，$\lim_{n\to\infty}\sum_{1\le j\le k(n)}E_{nj\varepsilon}=0$. 那么当 $n\to\infty$ 时，$\mathcal{L}(S_n)\to N(0,1)$.

证明 因为所有的 S_n 均值为 0，方差为 1，如同定理 9.5.6 的证明中那样，根据切比雪夫不等式可得，法则 $\mathcal{L}(S_n)$ 是一致紧的．因而由引理 9.5.5 知，只需证明 S_n 的特征函数逐点收敛到 $N(0,1)$ 的特征函数．根据独立性(定理 9.4.3)可得，

$$E\exp(itS_n)=\prod_{1\le j\le k(n)}E\exp(itX_{nj}).$$

下面 $\theta_r(r=1,2,\cdots,)$ 是复数且 $|\theta_r|\le 1$，它可能依赖于其他的变量．我们有下面的部分泰勒展开(见附录 B)：对所有的实数 u，

$$e^{iu}=1+iu+\theta_1u^2/2=1+iu-u^2/2+\theta_2|u|^3/6.$$

因为 $f(u):=e^{iu}$，故根据推论 B.4 可得 $|f^{(n)}(u)|=|i^ne^{iu}|\equiv 1$. 对所有的复数 z，当 $|z|\le 1/2$ 时，根据对数主分支展开的原理(见附录 B，推论 B.4 后的例(b))，

316 **9.6.2** $$\mathrm{plog}(1+z)=z+\theta_3 z^2.$$

取 $\varepsilon>0$，令 $\theta_4:=\theta_2(tx)$，$\theta_5:=\theta_1(tx)$，那么对每个 n 和 j，

$$E\exp(\mathrm{i}tX_{nj}) = \int 1+\mathrm{i}tx\mathrm{d}\mu_{nj}(x) + \int_{|x|\leqslant\varepsilon} -t^2x^2/2+\theta_4|tx|^3/6\mathrm{d}\mu_{nj}(x) + \int_{|x|>\varepsilon}\theta_5 x^2t^2/2\mathrm{d}\mu_{nj}(x).$$

因为 $EX_{nj}=0$，所以第一个积分为 1. 第二个积分中第一项的积分等于 $-t^2(\sigma_{nj}^2-E_{nj\varepsilon})/2$. 因为 $|x|\leqslant\varepsilon$ 时，$|x|^3\leqslant\varepsilon x^2$，所以第二项的绝对值最大为 $|t|^3\sigma_{nj}^2\varepsilon/6$. 最后一个积分的绝对值最大为 $t^2E_{nj\varepsilon}/2$. 因为 $0\leqslant E_{nj\varepsilon}\leqslant\sigma_{nj}^2$，$|\sigma_{nj}^2-E_{nj\varepsilon}(1-\theta_6)|\leqslant\sigma_{nj}^2-E_{nj\varepsilon}+E_{nj\varepsilon}=\sigma_{nj}^2$，所以可得

$$E\exp(itX_{nj}) = 1-t^2(\sigma_{nj}^2-E_{nj\varepsilon}(1-\theta_6))/2+\theta_7|t|^3\varepsilon\sigma_{nj}^2/6.$$

我们有 $\max_j\sigma_{nj}^2\leqslant\varepsilon^2+\sum_j E_{nj\varepsilon}$. 因此对任意实数 t，可选择足够小的 ε 和足够大的 n，使得对所有的 j，$z_{nj}:=E\exp(\mathrm{i}tX_{nj})-1$，且 $|z_{nj}|\leqslant1/2$. 然后应用式(9.6.2)，对每个固定的 t，令 $o(1)$ 表示收敛到 0 的任意项，当 $n\to\infty$ 时，

$$E\exp\left(\mathrm{i}t\sum_{j=1}^{k(n)}X_{nj}\right) = \exp(-t^2(1+o(1))/2+\theta_8|t|^3\varepsilon/6 + \theta_9\sum_{j=1}^{k(n)}[\theta_{10}t^2\sigma_{nj}^2(1+|t|\varepsilon/6)]^2).$$

现在

$$\sum_{j=1}^{k(n)}\sigma_{nj}^4\leqslant\left(\varepsilon^2+\sum_j E_{nj\varepsilon}\right)\sum_j\sigma_{nj}^2=\varepsilon^2+o(1),$$

将其代入前面的表达式得，

$$\exp(-t^2/2+o(1)+\theta_8\varepsilon|t|^3/6+\theta_{11}t^4[(\varepsilon^2+o(1))(1+|t|\varepsilon/6)^2]).$$

因为 $\varepsilon>0$ 是任意的，令 $\varepsilon\downarrow0$ 可得 $\exp(-t^2/2+o(1))$，因此当 $n\to\infty$ 时，$E\exp(\mathrm{i}tS_n)\to\exp(-t^2/2)$. □

在前面的定理 9.5.6 中已给出$\mathbb{R}$上独立同分布随机变量 X_1，X_2，…，且 $EX_1^2<\infty$ 的中心极限定理的证明，现在此定理的证明只是作为林德伯格定理应用的特例给出. 为简便起见，设 $EX_1^2=1$. 令 $k(n)=n$ 和 $X_{nj}=X_j/n^{1/2}$，则 $\sigma_{nj}^2=1/n$，$\sum_{1\leqslant j\leqslant n}\sigma_{nj}^2=1$，且对 $\varepsilon>0$， 317

$$E_{nj\varepsilon}:=\int_{|x|>\varepsilon}x^2\mathrm{d}\,\mathcal{L}(X_{nj})(x) = EX_{nj}^2 1_{\{|X_{nj}|>\varepsilon\}} = EX_1^2 1_{\{|X_1|>\varepsilon\sqrt{n}\}}/n,$$

因此由控制收敛定理，当 $n\to\infty$ 时，

$$\sum_{j=1}^n E_{nj\varepsilon} = EX_1^2 1_{\{|X_1|>\varepsilon\sqrt{n}\}}\to0.$$

有限测度 μ 的 n 阶卷积幂可定义为 $\mu^{*n}=\mu*\mu\cdots*\mu$，(共 n 个)，μ^{*0} 定义为 δ_0，其中对任意集合 A，$\delta_0(A):=1_{A(0)}$. 因此对概率测度 μ，μ^{*n} 是 $S_n=X_1+\cdots+X_n$ 的法则，其中 X_1，…，X_n 是有法则 μ 的独立同分布随机变量. 有限测度 μ 的指数定义为 $\exp(\mu):=\sum_{n\geqslant0}\mu^{*n}/n!$. 这些概念将在习题中进一步讨论.

对 $\lambda>0$，$\mathbb{N}$ 上的泊松分布 P_λ 定义为 $P_\lambda(k):=\mathrm{e}^{-\lambda}\lambda^k/k!$，$k=0$，1，…. 泊松分布是某三角形阵列的极限(习题 2)，尤其是二项分布(习题 3)的极限，并且已经应用于数据中，如放射性蜕变、染色体互换、细菌数和电话网中的突发事件等，见 Feller(1968，p. 159—164).

习题

1. 设 $X_{nk}=ks_k/n$，$k=1$，…，n，其中 S_k 是独立同分布随机变量且 $P(S_k=-1)=P(S_k=1)=1/2$. 求 σ_n 使得对变量 $X_{nk}/\sigma_n(k=1,\cdots,n)$ 能应用林德伯格定理.
2. 设 X_{nj} 是一个随机变量的三角形阵列，对每个 n，$j=1$，…，$k(n)$ 是独立的，且 $X_{nj}=1_{A(n,j)}$. 对某事件 $A(n,j)$，$P(A(n,j))=p_{nj}$. 令 S_n 为第 n 行的和，如果 $n\to\infty$ 时，$\max_j p_{nj}\to 0$，$\sum_j p_{nj}\to\lambda$，证明：$\mathcal{L}(S_n)\to P_\lambda$.
[提示：利用特征函数.]
3. 假定习题 2 的结果成立，证明：当 $n\to\infty$，$p=p_n\to 0$ 时，$np_n\to\lambda$，满足 $B_{n,p}(k):=\binom{n}{k}p^k(1-p)^{n-k}(k=0,1,\cdots,n)$ 的二项分布收敛到 P_λ.
4. 证明：对任意 $\lambda>0$，$e^{-\lambda}\exp(\lambda\delta_1)$ 是泊松法则 P_λ.（因此，泊松法则是一些简单非零测度的指数的乘积；另一方面，此指数给出了习题 2 和习题 3 的极限运算结果.）

318

5. 对 $\mathbb{R}$ 上任意有限(非负)测度 μ，证明：$e^{-\mu(\mathbb{R})}\exp(\mu)$ 是 $\mathbb{R}$ 上的概率测度. 并利用函数 $f(t)=\int e^{ixt}d\mu(x)$ 求出它的特征函数.
6. (卷积). 证明：对 $\mathbb{R}$ 上任意两个有限测度 μ 和 v，$\exp(\mu+v)=\exp(\mu)*\exp(v)$.
7. $\mathbb{R}$ 上的概率法则 P 称为*无穷可分的*(infinitely divisible)当且仅当对每个 $n=1$，2，…存在法则 P_n，其 n 阶卷积幂 $P_n^{*n}=P$，证明：任一正态法则是无穷可分的.
8. 对 $\mathbb{R}$ 上任意有限测度 μ，证明：习题 5 中的法则 $e^{-\mu(\mathbb{R})}\exp(\mu)$ 是无穷可分的.
9. $\mathbb{R}$ 上的法则 P 称为*稳定的*(stable)当且仅当对某些仿射函数 $A_n(x)=a_nx+b_n$ 和每个 $n=1$，2，…，$P_n^{*n}=P\circ A_n^{-1}$，证明：所有的正态法则是稳定的.
10. 证明：所有稳定的法则是无穷可分的.(9.8 节将介绍更多无穷可分和稳定的法则)
11. (排列). 对每个 $n=1$，2，…，一个*排列*(permutation)是从 $\{1,\cdots,n\}$ 映射到其自身的一个1－1 的函数. 对于一个给定的 n，所有这样的排列构成的集合称为*对称群*(symmetric group)，通常记为 S_n，这里记为 $\mathcal{S}_n$，它有 $n!$ 个元素. 设 μ_n 是其上的均匀法则，对每个排列 s，$\mu_n(\{s\})=1/n!$. 对每个 $s\in\mathcal{S}_n$，考虑序列 $a_1:=1$ 以及 $\forall j=2,3\cdots$ 当 $i<j$，$a_j\neq a_i$ 时，$a_j:=s(a_{j-1})$. 令 $J(1)$ 是使其成立的最大的 j.
    (a) 证明：$s(a_{J(1)})=1$. 数 a_1，…，$a_{J(1)}$ 形成了排列 s 的一个*循环*(cycle). 如果 $J(1)<n$，设 i 是 $\{1,\cdots,n\}/\{a_1,\cdots,a_{J(1)}\}$ 中的最小数，令 $a_{J(1)+1}=i$，重复应用 s 形成第二个以 $a_{J(2)}$ 结束的循环且 $s(a_{J(2)})=i$，重复这个过程，直到 a_1，…，a_n 被定义且对某个 k，$J(k)=n$. 如果对某个 m，$r=J(m)$(循环在第 r 步完成)，则令 $Y_{nr}(s)=1$，否则 $Y_{nr}(s)=0$.
    (b) 证明：对每个 n，$(\mathcal{S}_n,\mu_n)$ 上的随机变量 $Y_{nr}(r=1,\cdots,n)$ 是独立的.
    (c) 对每个 n 和 r，试求 EY_{nr} 和 $\mathrm{var}Y_{nr}$. 令 $T_n:=\sum_{1\leqslant r\leqslant n}Y_{nr}$，那么 $T_n(s)$ 是在排列 s 中不同循环的次数. 令 $\sigma(T_n):=(\mathrm{var}(T_n))^{1/2}$，$X_{nj}:=(Y_{nj}-EY_{nj})/\sigma(T_n)$.
    (d) 证明：对 X_{nj} 应用林德伯格定理，并求常数 a_n 和 b_n，使得当 $n\to\infty$ 时，$\mathcal{L}((T_n-a_n)/b_n)\to N(0,1)$，其中对某个 α，β，$a_n=(\log n)^\alpha$，$b_n=(\log n)^\beta$.

319

9.7 独立实值随机变量的和

一个实数级数收敛，若没有给出具体的和，则它的和是有限的. 一个收敛的随机变量的和是几乎必然有限的. 在给定意义下(如几乎必然、依概率或依法则)，随机变量和 $\sum_{n\geqslant 1}X_n$ 的收敛性意味着部分和序列 $S_n:=X_1+\cdots+X_n$ 在给定意义下的收敛性. 对于独立的随机变量和，我们有下面的定理.

9.7.1 莱维等价性定理　对一列独立实值随机变量 X_1，X_2，…，下面的命题是等价的：

(Ⅰ) $\sum_{n=1}^{\infty} X_n$ 几乎必然收敛.

(Ⅱ) $\sum_{n=1}^{\infty} X_n$ 依概率收敛.

(Ⅲ) $\sum_{n=1}^{\infty} X_n$ 依法则收敛，即当 $n\to\infty$ 时，$\mathcal{L}(S_n)$收敛到$\mathbb{R}$上的某个法则 μ.

如果这些条件不成立，那么 $\sum_{n=1}^{\infty} X_n$ 是几乎必然发散的.

注：如果 X_j 总是非负的，则 S_n 对 n 是非减的，在没有独立性这个条件时上面的三个命题的等价性仍成立(详细证明留作习题1). 因此仅当对不同的 j, X_j 有不同的符号时，独立性才是必需的.

证明 对一般的序列，(Ⅰ)⇒(Ⅱ)，(Ⅱ)⇒(Ⅲ)(定理9.2.1和命题9.3.5). 下面证明(Ⅲ)⇒(Ⅱ)：假定(Ⅲ)成立，根据命题9.3.4知，法则$\mathcal{L}(S_n)$是一致紧的. 因此对任意 $\varepsilon>0$，存在 $M<\infty$，使得对所有的 n，$P(|S_n|>M)<\varepsilon/2$. 那么对所有的 n 和 k，$P(|S_n-S_k|>2M)<\varepsilon$，因此，所有$\mathcal{L}(S_n-S_k)$的集合是一致紧的. 为证明(Ⅱ)，根据定理9.2.2和定理9.2.3，只需证明对樊畿度量 α，$\{S_n\}$是一个柯西序列. 如果不是，则存在 $\delta>0$ 和子序列 $n(i)$，使得对所有的 i，$\alpha(S_{n(i)}, S_{n(i+1)})>\delta$. 令 $Y_i:=S_{n(i+1)}-S_{n(i)}$，那么法则$\mathcal{L}(Y_i)$是一致紧的，所以由定理9.3.3可知，有子序列 $Q_j:=\mathcal{L}(Y_{i(j)})$收敛到某法则 Q. 注意，对所有的 i，$\alpha(Y_i, 0)>\delta$，因此 $P(|Y_i|>\delta)>\delta$. 取连续函数 f，满足$f(0)=0$，$0\leqslant f\leqslant 1$，并且对 $|x|>\delta$，$f(x)=1$. 那么 $Ef(Y_j)>\delta$，因此$Ef(Y_{i(j)})$不收敛到0，并且 $Q\neq\delta_0$.

设 $P_j:=\mathcal{L}(S_{n(i(j))})$，那么 $P_j\to\mu$，$P_j*Q_j\to\mu$，并由卷积的连续性(定理9.5.9)得，$P_j*Q_j\to\mu*Q$，因此(由引理9.3.2)$\mu=\mu*Q$，但这与命题9.5.10矛盾，故(Ⅲ)⇒(Ⅱ)的证明证毕.

320

在定理5.1.5后给出了范数的一般定义. 在(Ⅱ)⇒(Ⅰ)的证明中将用到下面关于随机变量的范数不等式. 它的叙述和证明将推广到无限维范数空间，但实际上在这里仅需要对通常的欧几里得范数$\|x\|=(x_1^2+\cdots+x_k^2)^{1/2}$的结果.

9.7.2 Ottaviani 不等式 设 $X_1, X_2, \cdots, X_n$ 是取值于$\mathbb{R}^k$的独立随机变量且 $S_j:=X_1+\cdots+X_j$，$j=1, \cdots, n$. 设$\|\cdot\|$是$\mathbb{R}^k$上的任意范数，假设对某个 $\alpha>0$ 和 $c<1$，$\max_{j\leqslant n} P\left(\|S_n-S_j\|>\alpha\right)\leqslant c$. 那么

$$P\left(\max_{j\leqslant n}\|S_j\|\geqslant 2\alpha\right)\leqslant(1-c)^{-1}P(\|S_n\|\geqslant\alpha).$$

证明 若存在使得$\|S_j\|\geqslant 2\alpha$ 的最小的 $j\leqslant n$，则令 $m:=m(\omega)$，若不存在这样的 j，则令 $m:=n+1$. 那么

$$P(\|S_n\|\geqslant\alpha)\geqslant P\left(\|S_n\|\geqslant\alpha, \max_{j\leqslant n}\|S_j\|\geqslant 2\alpha\right)=\sum_{j=1}^{n}P(\|S_n\|\geqslant\alpha, m=j)$$
$$\geqslant\sum_{j=1}^{n}P(\|S_n-S_j\|\leqslant\alpha, m=j).$$

因为 $m=j$ 蕴涵着$\|S_j\|\geqslant 2\alpha$，$\|S_n-S_j\|\leqslant\alpha$，则可推出$\|S_n\|\geqslant\alpha$. 因为$\|S_n-S_j\|$是 $X_{j+1}, \cdots, X_n$ 的函数，事件$\{m=j\}$是 $X_1, \cdots, X_j$ 的函数，所以$\|S_n-S_j\|$与事件$\{m=j\}$是独立的，因此最后的和等于

$$\sum_{j=1}^{n}P(\|S_n-S_j\|\leqslant\alpha)P(m=j)\geqslant(1-c)\sum_{j=1}^{n}P(m=j)$$

$$= (1-c)P\left(\max_{j\leqslant n}\|S_j\|\geqslant 2\alpha\right),$$

故不等式得证. □

现在假定(Ⅱ)成立，令 $c=1/2$. 考虑 $0<\varepsilon<1$，取足够大的 K，使得对一切 $n\geqslant K$，$P(|S_n-S_K|\geqslant\varepsilon/2)<\varepsilon/2$. 取 $\alpha=\varepsilon/2$，应用 Ottaviani 不等式于变量 X_{K+1}，X_{K+2}，…，则对任意 $n\geqslant K$，$P(\max_{K\leqslant j\leqslant h}\|S_j-S_K\|\geqslant\varepsilon)\leqslant\varepsilon$. 这里令 $n\to\infty$，则 $\varepsilon\downarrow 0$，$K\to\infty$，由引理 9.2.4 知，原级数几乎必然收敛，故(Ⅰ)得证，因此(Ⅰ)与(Ⅲ)是等价的.

最后，由科尔莫戈罗夫 0－1 律(8.4.4)知，$\sum_{n=1}^{\infty}X_n$ 或者几乎必然收敛或者几乎必然发散，这样就证得了定理 9.7.1. □　321

现在我们来探讨收敛的具体条件. 对任意实值随机变量 X，X 在 ± 1 处的截断函数定义为

$$X^1:=\begin{cases}X, & \text{若 } |X|\leqslant 1\\ 0, & \text{若 } |X|>1.\end{cases}$$

或回顾对任意实值随机变量 X，EX 有定义且有限，当且仅当 $E|X|<\infty$ 且方差 $\sigma^2(X)$ 是有定义的，如果 $EX^2<\infty$，$\sigma^2(X):=E((X-EX)^2)=EX^2-(EX)^2$. 下面有一个准则.

9.7.3 三级数定理　对独立实值随机变量 X_j 的级数(如定理 9.7.1 中所述)，其收敛性(几乎必然，或等价地依概率或依法则)与下面的条件是等价的：

(Ⅳ)下面的三个(实数的)级数是收敛的：

(a) $\sum_{n=1}^{\infty}P(|X_n|<1)$.

(b) $\sum_{n=1}^{\infty}EX_n^1$.

(c) $\sum_{n=1}^{\infty}\sigma^2(X_n^1)$.

注：在(b)中，绝对收敛——即 $\sum_n|EX_n^1|$ 的收敛性是不必要的. 例如，常数项级数 $\sum_n(-1)^n/n$ 满足(a)、(b)和(c).

证明　因为在定理 9.7.1 中已经证得(Ⅰ)和(Ⅲ)是等价的，在此只需证明(Ⅳ)⇒(Ⅱ)和(Ⅰ)⇒(Ⅳ)即可.

假定(Ⅳ)成立，那么由(a)和博雷尔－坎泰利引理(8.3.4)可知，当 $m\to\infty$ 时，P(对某个 $n\geqslant m$，$X_n\neq X_n^1$)$\to 0$. 因此，$\sum_n X_n$ 收敛当且仅当 $\sum_n X_n^1$ 收敛. 所以在证明依概率收敛的过程中可假定 $X_n=X_n^1$，即对所有的 n 和 ω，$|X_n(\omega)|\leqslant 1$. 然后由(b)可知，EX_n 的和是收敛的. 将所有的 X_n 除以 2(即 $|X_n|\leqslant 1/2$，$|EX_n|\leqslant 1/2$)不影响级数的收敛性. 因此 $\sum_n X_n$ 收敛当且仅当 $\sum_n X_n-EX_n$ 收敛. 取 X_n-EX_n，可假定对所有的 n，$EX_n=0$. 然后由级数(c)知，$\sum_n EX_n^2<\infty$，且对所有的 n 和 ω，$|X_n|\leqslant 1$.

对任意 $n\geqslant m$，由独立性(定理 8.1.2)得，

$$E((S_n-S_m)^2)=\sum_{m<j\leqslant n}EX_j^2\to 0, m\to\infty.$$

322

那么由切比雪夫不等式(8.3.1)可得，对任意 $\varepsilon>0$，$P(|S_n-S_m|>\varepsilon)\to 0$，因此，级数依概率收敛(利用定理 9.2.3)，故(Ⅱ)成立.

下面证明(Ⅰ)⇒(Ⅳ)，假定(Ⅰ)成立，那么由博雷尔-坎泰利引理和独立性可知，级数(a)收敛，再取 $|X_n|\leqslant 1$. 如果(c)发散，那么对每个 $m\leqslant n$，令 $S_{mn}:=\sum_{m\leqslant j\leqslant n}X_j$，选取足够大的 m，使得对所有的 $n\geqslant m$，$P(|S_{mn}|>1)<0.1$. 令

$$\sigma_{mn}:=(\mathrm{var}(S_{mn}))^{1/2}=\left(\sum_{m\leqslant j\leqslant n}\sigma^2(X_j)\right)^{1/2}\to\infty,n\to\infty.$$

令 $T_{mn}:=(S_{mn}-ES_{mn})/\sigma_{mn}$，或如果 $\sigma_{mn}=0$，令 $T_{mn}=0$，根据林德伯格定理(9.6.1)，对每个 m，当 $n\to\infty$ 时，$\mathcal{L}(T_{mn})\to N(0,1)$. 因为 $|X_j-EX_j|\leqslant 2$ 且方差的和是发散的，所以这里对足够大的 n，$E_{nj\varepsilon}=0$. 下面证明：因为 $\sum_{m\leqslant j\leqslant n}X_j$ 是具有大方差的渐近正态分布，因此，依概率不可能是小的. 对收敛的法则，分布函数收敛，其极限函数是连续的(定理 9.3.6)，因此，当 $n\to\infty$ 时，

$$P(S_{mn}\geqslant ES_{mn}+\sigma_{mn})=P(T_{mn}\geqslant 1)\to N(0,1)([1,\infty))>0.15>0.1.$$

然后选择 m，使得当 $n\to\infty$ 时，$ES_{mn}\leqslant-\sigma_{mn}+1\to-\infty$. 同理，

$$P(S_{mn}\leqslant ES_{mn}-\sigma_{mn})=P(T_{mn}\leqslant-1)\to N(0,1)((-\infty,-1])>0.15>0.1.$$

这蕴涵着 $ES_{mn}\to+\infty$，即得出矛盾，因此级数(c)收敛. 现在考虑级数 $\sum_n(X_n-EX_n)/2$，对此级数，三个级数都收敛. 如前面所证明的，因为(Ⅳ)⇒(Ⅱ)⇒(Ⅰ)，所以有 $\sum_n X_n-EX_n$ 几乎必然收敛. 那么减式 $\sum_n EX_n$ 是收敛的，因此级数(b)收敛，即(Ⅳ)得证，故三级数定理证毕. □

尽管上面的证明没有用到切比雪夫不等式，下面对切比雪夫不等式的改进(科尔莫戈罗夫不等式)是科尔莫戈罗夫对三级数定理的原始证明中的一部分.

9.7.4 科尔莫戈罗夫不等式　如果 $X_1,\cdots,X_n$ 是独立实值随机变量，均值为 0，$S_k:=X_1+\cdots+X_k$，则对任意 $\varepsilon>0$，

323

$$P\left(\max_{1\leqslant k\leqslant n}|S_k|>\varepsilon\right)\leqslant\varepsilon^{-2}\sum_{j=1}^{n}\sigma^2(X_j).$$

证明　如果 $ES_n^2=+\infty$，因为右边也等于 ∞，所以不等式成立，因此假设 $ES_n^2<\infty$. 令 $A_n=\{\max_{k\leqslant n}|S_k|>\varepsilon\}$，$B(k)=\{\max_{j<k}|S_j|\leqslant\varepsilon<|S_k|\}$，$k=1,\cdots,n$，那么 A_n 是不相交事件 $B(k)$ 的并. 由独立性可得，$ES_k1_{B(k)}(S_n-S_k)=0$ 且 $1_{B(k)}=1_{B(k)}^2$，因此，

$$\begin{aligned}\int_{B(k)}S_n^2\mathrm{d}P&=E((S_k1_{B(k)})^2)+E(|(S_n-S_k)1_{B(k)}|^2)\\&\geqslant ES_k^21_{B(k)}\geqslant\varepsilon^2P(B_k).\end{aligned}$$

从而

$$P(A_n)=\sum_{k=1}^{n}P(B(k))\leqslant\varepsilon^{-2}\sum_{k=1}^{n}\int_{B(k)}S_n^2\mathrm{d}P\leqslant\varepsilon^{-2}ES_n^2.$$ □

习题

1. 证明：定理 9.7.1 后的注对任意非负随机变量 X_j(不必独立)，级数 $\sum_j X_j$ 的三种收敛性是等价的.

2. 如果 X_1，X_2，…是独立随机变量，$\sum_n EX_n$ 是收敛的且 $\sum_n \sigma^2(X_n) < \infty$，证明：$\sum_n X_n$ 几乎必然收敛．[提示：不要截断函数(例如，不考虑 X_n^1)；应用而不是叙述三级数的部分证明和9.7.4.]
3. 设 $A(n)$ 是独立事件且满足 $P(A(n)) = 1/n$，试求常数列 c_n 在满足什么条件时，$\sum_n 1_{A(n)} - c_n$ 几乎必然收敛？
4. 设 X_1，X_2，…，是独立同分布实值随机变量且满足 $EX_1 = 0$ 和 $0 < EX_1^2 < \infty$，试求常数列 c_n 在满足什么条件时，$\sum_n c_n X_n$ 收敛？
5. 设 X_n 是独立同分布的且服从柯西分布 $P(X_n \in A) = \pi^{-1}\int_A \mathrm{d}x/(1+x^2)$，试求常数列 a_n 在什么条件时，$\sum_n a_n X_n$ 几乎必然收敛？可以利用柯西分布的特征函数为 $\mathrm{e}^{-|t|}$ 这一事实．
6. 设 X_j 是随机变量且满足对 $\forall j$，$EX_j^2 \leqslant M \leqslant \infty$ 和对 $i \neq j$，$EX_j = EX_iX_j = 0$，证明：X_j 满足强大数定律．[提示：
   (a)令 $s(n) := n^2$．证明 $S_{s(n)}/n^2 \to 0$ a. s..
   (b)令 $D(n) := \max\left\{ \left| S_k - S_{s(n)} \right| : n^2 \leqslant k < (n+1)^2 \right\}$．证明：$D_n/n^2 \to 0$ a. s..
   (c)结合(a)和(b)完成证明．]

324

7. 设 $\sum_j x_j$ 是一个收敛的实数级数且 b_n 是正数列，当 $n \to \infty$ 时，$b_n \uparrow +\infty$．如下所示证明 $\lim_{n\to\infty}\left(\sum_{1\leqslant k\leqslant n} b_k x_k \right)/b_n = 0$.
   (a)如果对所有的 k，$x_k \geqslant 0$，证明其服从控制收敛定理．
   (b)对一般情况的提示：
   $\sum_{1\leqslant k\leqslant n} b_k x_k = b_1(x_1 + \cdots + x_n) + (b_2 - b_1)(x_2 + \cdots + x_n) + \cdots + (b_n - b_{n-1})x_n$(一种部分和)，部分和 $\sum_{1\leqslant j\leqslant n} x_j$ 是绝对有界的，比如界为 M. 给定 $\varepsilon > 0$，选择足够大的 r，使得对所有的 $n \geqslant r$，$|x_r + \cdots + x_n| < \varepsilon/2$. 那么对使得 $Mb_r/b_n < \varepsilon/2$ 成立的足够大的 n，证明 $\left| \sum_{1\leqslant k\leqslant n} b_k x_k \right|/b_n < \varepsilon$.
8. 设 X_1，X_2，…，是独立实值随机变量且对所有的 k，$EX_k = 0$ 和 $EX_k^2 < \infty$．设 $0 < b_n \uparrow +\infty$，如果 $\sum_k EX_k^2/b_k^2 < \infty$，证明：$S_n/b_n \to 0$ a. s..[提示：由本节的结论，首先证明 $\sum_k X_k/b_k$ 几乎必然收敛，然后应用习题7的结论．]
9. 如果 X_1，X_2，…是独立同分布的且满足 $EX_1 = 0$，$EX_1^2 < \infty$，证明：$\forall \delta > 0$，$S_n/(n^{1/2}(\log n)^{0.5+\delta}) \to 0$ a. s..(用习题8的结论．)注意这是改进强大数定律(8.3.5)的一种方法，尽管是 EX_k^2 而不是 $E|X_k|$ 有限．(12.5节给出一个强化的 b_n，其中 $\log n$ 由 $\log\log n$ 替换．)

*9.8 莱维连续性定理：无穷可分法则及稳定法则

莱维(Lévy)连续性定理说的是，如果特征函数收敛(收敛到连续函数)，那么它所对应的法则也收敛．在法则序列是一致紧的假设下，此结论在引理9.5.5中已得到证明．但这里将证明特征函数收敛到连续的极限蕴涵着一致紧性．这个证明将以下面的结论为基础．

9.8.1 舍尾不等式 设 P 是 $\mathbb{R}$ 上的概率法则且特征函数为 $f(t) := \int \mathrm{e}^{\mathrm{i}xt}\mathrm{d}P(x)$，那么对任意满足 $0 < u < \infty$ 的 u，

$$P\left(|x| \geqslant \frac{1}{u}\right) \leqslant \frac{7}{u}\int_0^u 1 - \mathrm{Re}f(v)\,\mathrm{d}v.$$

证明 我们首先证明一个断言：对任意满足 $|t| \geqslant 1$ 的 t，$(\sin t)/t \leqslant \sin 1$.

325

因为 $(\sin t)/t$ 是偶函数，所以可假定 $t \geqslant 0$. 由于 $\sin 1 \approx 0.84 > 0.8$，因此对 $|t| \geqslant 1.3$ 断言成立；

而当 $1\leqslant t\leqslant 1.3$ 时，$(\sin t)/t$ 是递减的，断言得证．

然后由 Tonelli-Fubini 定理(4.4.5)得，

$$\frac{1}{u}\int_0^u\int_{-\infty}^{\infty}1-\cos(vx)\mathrm{d}P(x)\mathrm{d}v=\int_{-\infty}^{\infty}1-\frac{\sin(ux)}{ux}\mathrm{d}P(x)$$
$$\geqslant(1-\sin 1)P(|x|\geqslant 1/u).$$

由于 $\sin 1<6/7$，故 9.8.1 得证． □

9.8.2 莱维连续性定理 如果 P_n 是 $\mathbb{R}^k$ 上的法则，对所有的 t，其特征函数 $f_n(t)$ 收敛到某 $f(t)$，其中 f 沿着坐标轴在 0 点连续，那么对特征函数为 f 的某法则 P，$P_n\underset{\mathcal{L}}{\to}P$.

证明 考虑坐标轴 $t=(0,\cdots,t_j,0,\cdots,0)$，其中 $j=1,\cdots,k$，给定 $\varepsilon>0$，取足够小的 $\delta>0$，使得对坐标轴上的 t，当 $|t|\leqslant\delta$ 时，$|f(t)-1|<\varepsilon/(7k)$. 那么由(9.8.1)和控制收敛定理可得，对足够大的 n(如 $n\geqslant n_0$)和所有的 j，$P_n(|x_j|\geqslant 1/\delta)<\varepsilon/k$. 对每个 $n<n_0$，由可数可加性知，存在 $M_n<\infty$，使得对 $j=1,\cdots,k$，$P_n(|x_j|\geqslant M_n)<\varepsilon/k$. 令 $M:=\max(1/\delta,\max\{M_j:1\leqslant j<n_0\})$，那么对所有的 n 和 j，$P_n(|x_j|\geqslant M)<\varepsilon/k$. 如果 C 是立方体 $\{x:|x_j|\leqslant M,j=1,\cdots,k\}$，那么对所有的 n，$P_n(C)>1-\varepsilon$. 因此 P_n 是一致紧的．那么由引理 9.5.5 可得结论． □

定义 $\mathbb{R}$ 上的法则称为无穷可分的，如果对每个 $n=1,2,\cdots$，存在法则 P_n，使得其 n 次卷积幂 $P_n^{*n}=P$，即存在独立同分布．随机变量 $X_{n1},\cdots,X_{nn}$，使得 $\mathcal{L}(X_{n1}+\cdots+X_{nn})=P$.（对每个 j，令 $\mathcal{L}(X_{nj})=P_n$）

例如，因为对每个 n，$N(m,\sigma^2)=N(m/n,\sigma^2/n)^{*n}$，所以正态法则是无穷可分的．由特征函数的性质(乘性性质和唯一性)知，特征函数为 f 的法则 P 是无穷可分的，当且仅当对每个 $n=1$，2，…特征函数 f_n(关于某法则 P_n)使得对所有的 t，$f(t)=f_n(t)^n$. 在这种情况下，函数 f 称为无穷可分的．

下面将给出无穷可分法则和稳定法则的一些主要结论，大多数没有给出证明，更多细节可参考 Breiman(1968，第 9 章)和 Loève(1977，§23).

326

$\mathbb{N}$ 上参数为 λ 的泊松法则 P_λ，其中 $P_\lambda(k)=\mathrm{e}^{-\lambda}\lambda^k/k!$ $(k=0,1,\cdots)$是无穷可分的且对所有的 n，满足$(P_{\lambda/n})^{*n}=P_\lambda$. 其特征函数为 $\exp(\lambda(\mathrm{e}^{it}-1))$. 若随机变量 X 有无穷可分法则，那么对任意实数 b 和 c，$bX+c$ 也有无穷可分法则．因此 $\exp(\mathrm{i}ct+\lambda(\mathrm{e}^{ibt}-1))$是一无穷可分特征函数．进而，任意两个无穷可分特征函数的积也是无穷可分的．下面对 $\mathbb{R}$ 上的任意有限法则 μ，$g(t):=\exp(\mathrm{i}ct+\int\mathrm{e}^{itx}-1\mathrm{d}\mu(x))$ 是无穷可分特征函数(首先假设 μ 是单点测度，然后是一有限集中的法则；由莱维连续性定理再求极限).

反之，一个无穷可分特征函数 f 不可能为 0(见习题 6)，因此，$\log f$ 的唯一连续形式可由 $\log f(0)=0$ 来定义．那么对某法则 P_n 的特征函数 f_n，$f=f_n^n$. 由连续性知，$f_n\equiv\exp((\log f)/n)$，因此，当 $n\to\infty$ 时，$n(f_n-1)$ 收敛到 $\log f$，且

$$n(f_n-1)(t)=n\int_{-\infty}^{\infty}\mathrm{e}^{itx}-1\mathrm{d}P_n(x).$$

因此，f 必定是形如 g 的特征函数的极限．

事实上，μ 在 0 点附近有限的条件不是必需的，只要 $\mu\{|x|>1\}$ 是有限的且 $\int_{|x|<1}|x|\mathrm{d}\mu(x)<\infty$ 即可．那么 $g(t)$ 定义中的积分是收敛的且能给出一无穷可分的特征函数．同理，如果以适当的方式

取极限，在被积函数中 ict 这一项对于变量 c 可生成 $-itx$，$-itx$ 允许 μ 在 0 点附近满足更弱的条件 $\int_{|x|<1}|x|^2 d\mu(x) < \infty$. 那么 μ 上的这两个条件可以合并为 $d\mu(x) = (1+x)^2 dG(x)/x^2$，其中 G 为有限测度. 这将有下面的形式.

9.8.3 莱维–辛钦公式 *$\mathbb{R}$ 上的特征函数 f 是无穷可分的当且仅当它是形如 $e^{h(t)}$ 的，其中*

$$h(t) = ict + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}dG(x),$$

这里 $c \in \mathbb{R}$，G 是 $\mathbb{R}$ 上的有限非负测度，定义被积函数在 0 点处连续且值为 $-t^2/2$.

对正态测度 $N(\mu, \sigma^2)$ 的特征函数，我们有 $c = \mu$，且 G 是 σ^2 在 0 点的点式群体. [327]

特征函数为 f 的法则(或 f 本身)称为*稳定的*，当且仅当对所有的 $n = 1, 2, \cdots$，存在某个 a_n 和 b_n，使得对所有的 t，$f(t)^n = f(a_n t)\exp(ib_n t)$. (如同 9.6 节的习题 9). 对称稳定分布函数有下面的简单形式.

9.8.4 定理 *特征函数为 f 的对称法则 $P(dP(-x) \equiv dP(x))$ 是稳定的当且仅当对某个 $c \geqslant 0$，$0 < p \leqslant 2$ 及 $\forall t$，有 $f(t) = \exp(-c|t|^p)$.*

这里 p 称为稳定法则的指数，在退化情况下，$c = 0$，$f \equiv 1$ 且法则是点质量. 在 $p = 2$，$c = \sigma^2/2$ 时，对称正态法则 $N(0, \sigma^2)$ 是稳定的. 如果在稳定性的定义中取 $a_n = n^{1/p}$，则非对称稳定法则的指数为 p. 每个稳定的法则都有指数 p，且 $0 < p \leqslant 2$. 指数为 2 的稳定法则仅有正态法则. 具有指数 $p < 2$ 的稳定法则的随机变量 X，对 $r > 0$，满足 $E|X|^r < \infty$ 当且仅当 $r < p$. 例如，密度为 $1/(\pi(1+x^2))$ 的柯西分布是指数为 1 的稳定分布，其特征函数是 $e^{-|t|}$. 对大多数指数的值，对称稳定密度是不知道的. 对一般的(可能是非对称的)稳定特征函数，我们有下面的结论.

9.8.5 定理 *特征函数 f 是稳定的当且仅当或者它是正态的或者它的稳定指数为 p，$0 < p < 2$，并且 $f(t) \equiv e^{h(t)}$，这里对某个 $m \in \mathbb{R}$，$c > 0$，$|\beta| \leqslant 1$，*

$$h(t) \equiv imt - c|t|^p(1 - i\beta\,\mathrm{sgn}(t)\omega(t,p)),$$

其中 $p \neq 1$ 时，$\omega(t, p) = \tan(\pi p/2)$；$p = 1$ 时，$\omega(t, p) = -(2/\pi)\log|t|$；如果 $t > 0$，$\mathrm{sgn}(t) = 1$；如果 $t = 0$，$\mathrm{sgn}(t) = 0$；如果 $t < 0$，$\mathrm{sgn}(t) = -1$.

结合前面关于无穷可分法则和稳定法则的结论，这个定理在 Loéve(1977，§23)中得到证明. 也可参考另外的文献《caveat lector》(Hall，1981).

习题

1. 证明：对所有的 $n = 1, 2, \cdots$ 函数 $\cos^{2n}(t)$ 是特征函数，并有当 $n \to \infty$ 时，对所有的 t，它收敛到函数 $f(t)$，但 $f(t)$ 不是特征函数，且其对应的法则不收敛.
2. 证明：对任意 $|t| \geqslant 1$，$(1 - \cos t)/t^2 \leqslant 1 - \cos 1$.
3. 对特征函数唯一性定理(9.5.1)另一种证明，令 $|x|^2 := (x, x)$，$x \in \mathbb{R}^k$，f_P 是 P 的特征函数. [328]
   (a) 证明：对任意 $t \in \mathbb{R}^k$，$c > 0$，对某个常数 $A(c, k)$，
   $$\int f_P(t-u)\exp(-c|u|^2/2)\,du = A(c,k) \times \int \exp(-|x|^2/2c) + i(x,t)\,dP(x).$$
   (b) 设 $\mathcal{F}$ 是所有由常数和形如 $x \mapsto \exp(i(x, t) - b|x|^2)$ $(b > 0, t \in \mathbb{R}^k)$ 的函数的复线性组合的集合. 如果对法则 P 和 Q，$f_P \equiv f_Q$，证明：对所有的 $g \in \mathcal{F}$，$\int g\,dP = \int g\,dQ$.
   (c) 取 $\mathbb{R}^k$ 上的一点紧化 K(定理 2.8.1)，并证明 $\mathcal{F}$ 中的所有函数都能扩张为 K 上的连续函数. 对 K 上的任意

连续函数 g，对结论 $\int g\mathrm{d}P=\int g\mathrm{d}Q$ 应用斯通-魏尔斯特拉斯定理，然后对任意 $g\in C_b(\mathbb{R}^k)$，由控制收敛定理可得到相同的结论，因此由引理 9.3.2 可得 $P=Q$.

4. 对不用舍尾不等式的莱维连续性定理的另一种证明，假定法则 P_n 的特征函数收敛到法则 P 的特征函数 f_P（如同对中心极限定理是成立的），令 $P(n):=P_n$，那么根据控制收敛定理可知，对所有的 $t\in\mathbb{R}^k$，当 $n\to\infty$ 时，

$$\int f_{P(n)}(t-u)\exp(-c|u|^2)\mathrm{d}u\to\int f_P(t-u)\exp(-c|u|^2)\mathrm{d}u.$$

和习题 3 证明相同，对 $\mathbb{R}^k$ 上任意满足在某紧集 D 外 $g=0$ 的连续函数 g，$\int g\mathrm{d}P_n\to\int g\mathrm{d}P$. 对每个 $\varepsilon>0$，在 $P(C)>1-\varepsilon$ 的集合 C 上，取 $g=1$. 应用引理 9.5.5 和定理 9.5.1 可得 P_n 是一致紧的．

5. 证明：对 $\mathbb{R}$ 上特征函数为 f 的任意法则 P 和 $u>0$，

$$\int_{|x|<1/u}x^2\mathrm{d}P(x)\leqslant 3u^{-2}(1-\mathrm{Re}f(u)).$$

[提示：证明对任意实数 t，在 $t=0$ 处利用导数有 $1-\cos(t)\geqslant t^2/2-t^4/24$. 然后证明

$$\int_{-\infty}^{\infty}1-\cos(ux)\mathrm{d}P(x)\geqslant\int_{|x|<1/u}\frac{1}{2}u^2x^2\left(1-\frac{u^2x^2}{12}\right)\mathrm{d}P(x).$$

右边被积函数的最后一部分是有界的．]

6. 如果 f 是一无穷可分特征函数，证明：对所有的 t，$f(t)\neq0$.［提示：令 f_n 是特征函数且 $f_n^n\equiv f$，证明 f 和 f_n 的复共轭是特征函数，因此 $|f|^2$ 和 $|f_n|^2$ 也是．那么根据 $f(t)$ 是否为 0 可得 $|f_n|^2(t)=|f|^{2/n}(t)\to H(t)$ 为 0 或 1，然后证明 H 是特征函数，因此，由连续性得 $H\equiv1$（定理 9.4.4）.］

7. 设 f 是 $\mathbb{R}$ 上的连续函数且对所有的 t，满足 $f(0)=1$，$f(t)=f(-t)$，并且使得限制于 $[0,\infty)$ 的 f 是凸的，且当 $t\to\infty$ 时，$f(t)\to0$. 证明：f 是一特征函数．［提示：首先假设 f 的图形是由有限多条直线段组成．这些直线段在 $[0,\infty)$ 有非负递增的斜率且一直递增到 t 轴．扩张右边的线段到 $t=0$，利用 $t<0$ 的镜像来定义一个新的函数 g. 考虑 $f-g$ 并且对直线段的数目用归纳法，然后由分段线性函数来逼近普通函数 f.］

329

8. (a)对任意 $p(0<p\leqslant1)$，由习题 7 的结果证明：$\exp(-|t|^p)$ 是特征函数．（如果 $0<p\leqslant2$，这也是正确的）
(b)可由定义直接证明(a)中的特征函数是稳定律的特征函数．

9. 证明：对 $p>2$，$\exp(-|t|^p)$ 不是特征函数．［提示：见 9.4 节习题 8(b).］

注释

9.1 节 第一个提出分布函数概念的人是谁？这个概念的发展经历了几个阶段．累积概率（例如，n 次独立试验有 k 次成功的概率）出现在伯努利等人的著作中．Laplace(1982)在他的书中和早期的工作中将连续分布函数作为密度函数的积分来处理，可能他是最早用这种方法的人．

9.2 节 对任意有限测度空间 $(\Omega,\mathcal{S},\mu)$ 和可分度量空间 (S,d)，恰好当 μ 是概率测度时，依测度收敛定义在 $L^0(\Omega,S)$ 上．F. Riesz(1909)对实值函数定义了依测度收敛并且注明（如勒贝格一样）逐点(a.e.)收敛序列依测度收敛．Riesz 证明了任意依测度收敛的序列都有几乎处处收敛的子序列．

Fréchet(1921, p.199—200)证明了依测度收敛是可度量的，其中度量为

$$\varphi(X,Y):=\inf\{\varepsilon+\mu\{\omega:d(X,Y)>\varepsilon\}:\varepsilon>0\}.$$

也可见 Fréchet(1928, p.90—91; 1937, p.191). Ky Fan(1944)定义了樊畿度量．他于 1914 年出生在中国杭州，1936 年在北京获得学士学位，1941 年在巴黎获得博士学位，且一直在巴黎生活到 1945 年，其后移居美国．

9.3 节 法则的收敛性，从一维分布函数的收敛性入手是一种比较经典的方法．$\mathbb{R}^2$ 上法则的收敛

性最初也是借助于二维分布函数 $P_n((-\infty, x]\times(-\infty, y])$ 来研究的(Khinchin, 1933, p. 11—16). 后来 A. D. Alexandrov(1940, 1941, 1943)给出了法则收敛的一般理论. 他的论文用很长的篇幅证明了一些基本的结论. 他在凸曲面几何方面的研究也很有名气. 如果想了解亚历山德罗夫(Alexandrov)的成果, 可参考 Efimov et al. (1962).

9.4 节 A. de Moivre(1733)发现了正态分布和中心极限定理之间的联系, 对于二项分布, 其中每个独立变量 X_j 有两种可能的取值 1 和 0, 所对应的概率分别为 p 和 $1-p$, 棣莫弗(De Moivre)更关注 $p=1/2$ 的情形, 但是在他早期的关于这方面的文献中对一般的 p 也进行了讨论, 这可从 de Moivre(1733)文章的题目(用 $(a+b)^n$ 而不是 $(1+1)^n$)看出. 1733 年, 已经 66 岁的棣莫弗除了代课和做顾问外, 没有担任过任何学术职务及取得任何学术位置, (Schneider, 1968—1969). Daw 和 E. S. Pearson(1972)写到"关于 1733 年其成果的记注共有 6 个版本, 其中的 5 个版本是与 Miscellanea Analytica(de Moivre. 1730)合订在一起", 也可参考 Archibald(1926). De Moivre 还把他 1733 年的成果翻译成英语并且将它写进了他的书《The Doctrine of Chance》的第 2 版(1738)中; De Moivre(1756, repr. 1967)是第 3 版也是最后一版, 并且是最容易得到的版本. 在那本书的 250 页可以看到对 $p\neq 1/2$ 的讨论. Todhunter(1865, p. 192—193)仅仅注意到对 $p=1/2$ 的讨论, 因此把对 $p\neq 1/2$ 的情况的研究归功于拉普拉斯是一个误导. Pearson(1924)指出棣莫弗讨论了一般的 p. 许多概率学家仍旧把正态分布称为高斯分布, 据 Pearson 的观点, 这是因为在 Laplace(1812)这本书出版的前些年高斯就公布了这些分布, 但是正态分布是在 18 世纪 70 年代出现在拉普拉斯的论文中, 并且比棣莫弗的研究要早 40 年, 对那个问题, 大多数对 $p=1/2$ 的情况进行处理. 另一方面, 记号 π(3.14…)和 e(2.71828…)应归功于欧拉(Euler), 而不是棣莫弗, 他用了很绕口的表达. 跟 Pearson 一样, 大多数统计学家愿用正态分布而不愿用高斯分布.

330

一个图书馆的目录单上至少有 19 本由 Todhunter 写的书, Todhunter(1865)是概率或统计学家. Baron(1976)说"许多孩子从上小学到读大学所学的数学课本都是 Todhunter 的教科书……Todhunter 最初不是数学家". Todhunter(1865, p. 193)确实提到了棣莫弗, "毫无疑问, 除了拉普拉斯外, Todhunter 是对概率论贡献最大的数学家," 作为棣莫弗的三大基本贡献(仅对 $p=1/2$)之一的渐近标准二项分布, Todhunter 把它称为由斯特林定理支持的伯努利定理的值的棣莫弗扩张. [Jakob]伯努利定理是二项分布的大数定律. 公式 $n!\sim(n/e)^n(2\pi n)^{1/2}(n\to\infty)$ 仅仅由(Stirling, 1730)发现, 当时棣莫弗自己也近乎发现了这个公式, 并应用了它. 斯特林利用这个公式算出了 $(2\pi)^{1/2}$ 的值, 对此, 棣莫弗首先给出了近似值, 最终给出了一个发散的表达; 在了解斯特林常数后, 棣莫弗(1730)用另一种方法证明了这个公式.

棣莫弗写道, 实际上, 根据正态密度函数在两点取值的比率, Laplace(1774)对正态密度给出了一个更清晰的处理并且证明了命题 9.4.1, 见 Stigler(1986, p. 360).

概率中特征函数的使用至少能追溯到 Cauchy(1853). 它们在现代理论中的重要性大部分体现在 P. Lévy(1925)的成果中. 在 Levy's Oeuvres(1973—1980)中概率论文从第 3 卷开始.

9.5 节 在棣莫弗阐述了中心极限定理的形式之后, 对服从二项分布的变量, Laplace(1812)实质性地扩展了这个定理. Lyapunoff(1901)对均值为 0、$E|X_i|^{2+\delta}<\infty$(某 $\delta>0$)的独立实值变量 X_j, 证明了这个定理. 此变量可能是同分布的或是更一般的, 这些假设比以前的假设更弱了. Lindeberg(1922a)首先证明了在 9.6 节给出的 $\mathbb{R}$ 上的中心极限定理, 其中仅假设 $EX_1=0$, $EX_1^2<\infty$. 这个定理及其证明参见 Lindeberg(1922a, 1922b)和 Lévy(1922a, 1922b, 1922c). 林德伯格的证明没有使用特

331

征函数．莱维发现了用特征函数证明的方法，在纠正了一些错误后，其证明比林德伯格的证明简短得多．Lévy(1922c)首先给出了唯一性定理(9.5.1)和连续性定理(9.5.5)的一种形式．林德伯格和莱维在证明唯一性定理的过程中都用了增加一个小的独立正态变量 Y 的方法．

在某种意义上讲，Pearson(1900)用了$\mathbb{R}^k$ 上的中心极限定理．显然，S. Bernstein(1927，p. 44—45)首先对 $k=2$ 给出了一般的严格证明．Khinchin(1933，p. 11—16)对 $k=2$ 作出了处理并声明对一般情况也成立．伯恩斯坦(Bernstein)和辛钦(Khinchin)的结果包含了一些不是同分布的和/或不独立的变量．在一个8页的伯恩斯坦数学断言(Alexandrov. et al.，1969)里涉及了一些遗传学方面的研究，但是在概率研究方面仅涉及了一小部分．伯恩斯坦在他的论文里解决了希尔伯特的第19个问题，分析了解决正则解析变量的问题．伯恩斯坦的最著名的工作是用多项式来逼近连续函数．

9.6节　林德伯格定理(9.6.1)有一个逆命题：如果当 $n\to\infty$ 时，$\max_j \sigma_{nj}^2\to 0$，$\mathcal{L}(S_n)\to N(0, 1)$，$EX_{nj}=0$，且 $\sum_j \sigma_{nj}^2 = 1$，那么 $\lim_{n\to\infty} = \sum_j E_{nj\varepsilon} = 0$. 因而也许更吃惊的是后边的条件是严格的，分别由Feller(1935)和 Lévy(1935，p. 386—388)独立地证明了．

Lindeberg(1922a，Satz Ⅲ)证明了他的中心极限定理(9.6.1)，形式略微有不同，因此，他第一次对独立同分布实值变量 X_1，X_2，…在简单条件 $EX_1=0$，$EX_1^2<\infty$ 下证明中心极限定理．

林德伯格和莱维在他们1922年的论文中讨论了三角形阵列，实际上不是正式的．Lévy(1970，p. 76)，在他的自传中，评论了他1922年的工作．Schweder(1980，p. 120，127)给出了简要介绍了林德伯格．见 Cramér(1976，p. 514)．

以 Poisson(1837)命名的分布可见 de Moivre(1711—1712)，还可参考 Hald(1984)．泊松生于1781年，卒于1840年．Sheynin(1977)介绍了泊松在概率方面的成果．

Gnedenko 和 Kolmogorov (1949)是关于独立变量和的极限定理的经典的专著．Gikhman、Kolmogorov、Korolyuk(1962)是为纪念格涅坚科(Gnedenko)的研究成果而写．

9.7节　三级数定理产生于 Khinchin 和 Kolmogorov(1925)的研究，后来由 Kolmogorov(1928，1929)和 Lévy(1937，第6章)证明了它的等价性定理．Ottaviani(1939)证明了其不等式．习题7的结果归功于克罗内克(Kronecker)，习题8的结果归功于科尔莫戈罗夫．

9.8节　舍尾不等式见 Loève(1977，p. 209)，其涉及 Lévy(1937，1954)和 Lévy(1922c)对 $k=1$ 时连续性定理的证明，根据 Bochner(1932，p. 72，223—224)可知，极限函数是处处连续的．各种扩张(到可数维，仅允许几乎处处收敛，极限函数仅仅在坐标轴上是连续的)不是很困难．它们的先后顺序这里没有给出．

无穷可分法则和稳定法则的理论是由 Lévy(1924，1925，1937)、Khinchin 和 Lévy(1936)发展起来的．Hall(1981)批判地回顾了这一理论的发展和稍后的一些结果．霍尔(Hall)强调了在证明过程中 β 符号的错误(在任何情况下，它都不影响结果的正确性)．

332

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到．

Alexandrov, A. D. (1940, 1941, 1943). Additive set-functions in abstract spaces. *Mat. Sbornik* (N.S.) *8*: 307–348; *9*: 563–628; *13*: 169–238.

Alexandrov, P. S., N. I. Akhiezer, B. V. Gnedenko, and A. N. Kolmogorov (1969). Sergei Natanovich Bernstein: Obituary. *Russian Math. Surveys 24*, no. 3, pp. 169–176. Transl. from *Uspekhi Mat. Nauk 24*, no. 3, pp. 211–218.

Archibald, R. C. (1926). A rare pamphlet of Moivre and some of his discoveries. *Isis 8*: 671–676.

Baron, Margaret E. (1976). Todhunter, Isaac. *Dictionary of Scientific Biography, 13*, 426–428.

Bernstein, Serge [Sergei Natanovich] (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann. 97*: 1–59.

Bochner, Salomon (1932). *Vorlesungen über Fouriersche Integrale*. Akademische Verlagsges. M.B.H., Leipzig. Repr. Chelsea, New York, 1948.

Breiman, Leo (1968). *Probability*. Addison-Wesley, Reading, Mass.

Cauchy, Augustin-Louis (1853). Sur les resultats moyens d'observations de même nature, et sur les resultats les plus probables. *Comptes Rendus Acad. Sci. Paris 37*: 198–206. Sur la probabilité des erreurs qui affectent des resultats moyens d'observations de même nature. Ibid., pp. 264–272. Mémoire sur les resultats moyens d'un très-grand nombre d'observations [Summary]. Ibid., pp. 381–385.

Cramér, Harald (1976). Half a century with probability theory: Some personal recollections. *Ann. Probability 4*: 509–546.

Daw, R. H., and E. S. Pearson (1972). Studies in the history of probability and statistics, XXX. Abraham De Moivre's 1733 derivation of the normal curve: a bibliographical note. *Biometrika 59*: 677–680.

Efimov, N. V., V. A. Zalgaller, and A. V. Pogorelov (1962). Alexander Danilovich Alexandrov (On his 50th birthday). *Russian Math. Surveys 17*, no. 6: 127–133. Transl. from *Uspekhi Mat. Nauk 17*, no. 6: 171–184.

Fan, Ky (1944). Entfernung zweier zufälligen Grössen und die Konvergenz nach Wahrscheinlichkeit. *Math. Zeitschr. 49*: 681–683.

Feller, Willy (1935). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Zeitschr. 40*: 521–559.

——— (1968). *An Introduction to Probability Theory and Its Applications*. Vol. 1, 3d ed. Wiley, New York.

Fréchet, Maurice (1921). Sur divers modes de convergence d'une suite de fonctions d'une variable. *Bull. Calcutta Math. Soc. 11*: 187–206.

——— (1928). *Les espaces abstraits* ... Gauthier-Villars, Paris.

——— (1937). *Recherches théoriques modernes sur la théorie des probabilités I*. Paris.

Gikhman, I. I., A. N. Kolmogorov, and V. S. Korolyuk (1962). Boris Vladimirovich Gnedenko (On his 50th birthday). *Russian Math. Surveys 17*, no. 4: 105–113. Transl. from *Uspekhi Mat. Nauk 17*, no. 4: 191–200.

Gnedenko, Boris V., and Andrei N. Kolmogorov (1949). *Limit Distributions for Sums of Independent Random Variables*. Translated, annotated, and revised by Kai Lai Chung. Addison-Wesley, Reading, Mass., 1954, 2d ed. 1968, with appendices by J. L. Doob and P. L. Hsu.

Hald, Anders (1984). A. de Moivre: "De Mensura Sortis" or "On the measurement of chance." Transl. with commentary. *Internat. Statist. Rev. 52*: 229–262.

Hall, Peter (1981). A comedy of errors: The canonical form for a stable characteristic function. *Bull. London Math. Soc. 13*: 23–27.

Hoffman, Kenneth, and Ray Kunze (1971). *Linear Algebra*. 2d ed. Prentice-Hall, Englewood Cliffs, N.J.

Khinchin, A. Ya. [Aleksandr Yakovlevich] (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*. Springer, Berlin.

———, and A. N. Kolmogoroff (1925). Über Konvergenz von Reihen deren Glieder durch den Zufall bestimmt werden. *Mat. Sbornik (Recueil Math.) 32*: 668–676. Russ. Summary, p. 677.

———, and P. Lévy (1936). Sur les lois stables. *C. R. Acad. Sci. Paris 202*: 374–376.

Kolmogoroff, A. N. [Kolmogorov, Andrei Nikolaevich] (1928, 1929). Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Math. Annalen 99*: 309–319. Bemerkungen zu meiner Arbeit . . . Ibid. *102*: 484–488.

Laplace, Pierre Simon de (1774). Mémoire sur la probabilité des causes par les événemens. *Mémoires de mathématique et de physique présentés a l'Académie royale des sciences, par divers savans & lus dans ses assemblées 6*: 621–656. Repr. in *Oeuvres complètes de Laplace, 8*, 325–366. Transl. into English and included (except an "unrelated" §7, on differential equations) in Stigler (1986, pp. 364–378).

——— (1812). *Théorie analytique des probabilités.* 3d ed. (1820). Mme Ve Courcier, Paris. Repr. in *Oeuvres complètes de Laplace,* 7.

——— (1886–1912). *Oeuvres complètes de Laplace.* Gauthier-Villars, Paris.

Lévy, Paul (1922a). Sur le rôle de la loi de Gauss dans la théorie des erreurs. *C. R. Acad. Sci. Paris 174*: 855–857.

——— (1922b). Sur la loi de Gauss. *C. R. Acad. Sci. Paris 174*: 1682–1684.

——— (1922c). Sur la détermination des lois de probabilité par leurs fonctions caractéristiques. *C. R. Acad. Sci. Paris 175*: 854–856.

——— (1924). Théorie des erreurs. La loi de Gauss et les lois exceptionelles. *Bull. Soc. Math. France 52*: 49–85.

——— (1925). *Calcul des probabilités.* Gauthier-Villars, Paris.

——— (1935). Propriétés asymptotiques des sommes de variables indépendantes ou enchaînées. *J. math. pures appl.* (Ser. 9) *14*: 347–402.

——— (1937). *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris. 2d ed. 1954.

——— (1970). *Quelques aspects de la pensée d'un mathématicien.* Blanchard, Paris.

——— (1973–1980). *Oeuvres de Paul Lévy.* 6 vols. Gauthier-Villars, Paris.

Lindeberg, J. W. (1922a). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Zeitschr. 15*: 211–225.

——— (1922b). Sur la loi de Gauss. *C. R. Acad. Sci. Paris 174*: 1400–1402.

Loève, Michel (1977). *Probability Theory I.* 4th ed. Springer, New York.

*Lyapunoff, A. M. (1901). Nouvelle forme du théorème sur la limite des probabilités. *Mem. Acad. Sci. St. Petersburg 12*, no. 5: 1–24.

de Moivre, Abraham (1711–1712). De Mensura Sortis seu, de Probabilitate Eventuum in Ludis a Casu Fortuito Pendentibus. *Phil. Trans. Roy. Soc. London 27*, whole no. 329. Transl. in Hald (1984).

——— (1730). *Miscellanea Analytica.* London.

——— (1733). *Approximatio ad Summam Terminorum Binomii* $(a + b)^n$ *in Seriem Expansi.* 7 pp. London. Repr. in Archibald (1926).

——— (1756, posth.). *The Doctrine of Chances*, 3d. ed. London. Repr. F. Case, London, and Chelsea, New York, 1967.

Ottaviani, G. (1939). Sulla teoria astratta del calcolo delle probabilità proposita dal Cantelli. *Giorn. Ist. Ital. Attuari 10*: 10–40.

Pearson, Karl (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.* (Ser. 5) *50*: 157–175.

——— (1924). Historical note on the origin of the normal curve of errors. *Biometrika 16*: 402–404.

Poisson, Siméon D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile précédés des règles générales du calcul des probabilités.* Bachelier, Paris.

Riesz, F. (1909). Sur les suites de fonctions mesurables. *Comptes Rendus Acad. Sci. Paris 148*: 1303–1305.

Schneider, Ivo (1968–1969). Der Mathematiker Abraham de Moivre, 1667–1754. *Arch. Hist. Exact. Sci. 5*: 177–317.

Schweder, Tore (1980). Scandinavian statistics, some early lines of development. *Scand. J. Statist. 7*: 113–129.

Sheynin, O. B. (1977). S. D. Poisson's work in probability. *Arch. Hist. Exact Sci. 18*: 245–300.

Stigler, Stephen M. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science 1*: 359–378.

Stirling, James (1730). *Methodus Differentialis*. London.

Todhunter, Isaac (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to That of Laplace*. Macmillan, London.



第10章　条件期望和鞅

条件期望为大量的概率计算提供了有用工具．鞅是与条件期望相关的随机变量序列，其满足某些极限定理，其中最大的作用是减弱了对随机变量独立性的要求．

10.1　条件期望

给定概率空间$(\Omega, \mathcal{S}, P)$和$\mathcal{S}$中的两个事件A和B，且有$P(B)>0$，A关于B的条件概率(conditional probability)定义为$P(A \mid B) := P(A\cap B)/P(B)$，这里$P(A\mid B)$读作“$A$关于$B$的条件概率”，“条件”两字可以被省略，因为“关于”已足可以表达“条件概率”的思想．例如，掷均匀硬币三次，假定第一次是正面，则至少有两次是正面的概率是3/4.

随机变量X关于事件B的条件期望(conditional expectation)(当它存在时)定义为

$$E(X\mid B) := \left(\int_B X\mathrm{d}P\right)\Big/ P(B).$$

然而更重要的是X关于$\mathcal{S}$中子σ-代数$\mathcal{A}$的条件期望，它定义如下：一个关于$\mathcal{A}$是可测的随机变量Y，使得对所有的$A\in\mathcal{A}$，$\int_A Y\mathrm{d}P = \int_A X\mathrm{d}P$，如果这样的$Y$存在，则记作$Y := E(X\mid \mathcal{A})$.

如果$\mathcal{A}=\mathcal{S}$，或更一般地，如果X关于$\mathcal{A}$是可测的，那么X自身满足$E(X\mid\mathcal{A})$的定义．更有趣的是，条件期望的非平凡情况就是X关于$\mathcal{A}$不可测．

336

例：

(i)将一个均匀的硬币掷两次，有一个包含四种不同结果的集合Ω：HH，HT，TH，TT，每个结果的概率为1/4．令X为结果是正面的次数：0，1，2．设$\mathcal{A}$是由第一次结果决定的σ-代数，则$\mathcal{A}=\{\varnothing, \{HH, HT\}, \{TH, TT\}, \Omega\}$．如果第一次是正面，那么$E(X\mid\mathcal{A})=3/2$，否则$E(X\mid\mathcal{A})=1/2$.

(ii)设X和Y是概率空间$(\Omega, \mathcal{S}, P)$上的独立随机变量，设$\mathcal{A}$是Y在其上是可测的最小的子σ-代数．因为对于Y的值域内的可测集B，$\mathcal{A}$是由所有$A=Y^{-1}(B)$的集合组成的，所以$E(X\mid\mathcal{A})=EX$．对任意这样的A，1_A和X是相互独立的且$E(1_A X)=E(1_A)EX$，因此，在这种情况下，条件期望变为一般的期望．

令$\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{S}, P)$，如果条件期望定义中的所有积分是有定义的且是有限的，那么取$A=\Omega$，$X\in\mathcal{L}^1$是必需的．这也是条件期望存在的充分条件.

10.1.1 定理　*对任意$X\in\mathcal{L}^1$，任意$\mathcal{S}$的子σ-代数$\mathcal{A}$，条件期望$E(X\mid\mathcal{A})$存在，且X关于$\mathcal{A}$的任意两个条件期望Z和Y几乎必然相等．*

证明　对所有的$A\in\mathcal{A}$，令$\mu(A) := \int_A X\mathrm{d}P$，那么$\mu$是$\mathcal{A}$上的可数可加符号测度，且关于$P$(限定到$\mathcal{A}$上)是绝对连续的，因此由Radon-Nikodym定理(推论5.6.2)知，对符号测度，存在$Y\in\mathcal{L}^1(\Omega, \mathcal{S}, P)$，使得对所有的$A\in\mathcal{A}$，$\mu(A) := \int_A Y\mathrm{d}P$．这样的$Y$满足$E(X\mid\mathcal{A})$的定义．如果$Z$也满足其定义，那么对所有的$A\in\mathcal{A}$，$\int_A Y-Z\mathrm{d}P=0$．由于$Y$和$Z$是$\mathcal{A}$-可测的，故事件$\{Y>Z\} := \{\omega: Y(\omega)>Z(\omega)\}$和$\{Y<Z\}\in\mathcal{A}$．在这些集合上积分可得，对$P$(如果$P\{Y>Z\}>0$，那么对某个

n，$P\{Y>Z+1/n\}>0$），它们的概率为 0，因此 $Y=Z$ a. s.. □

期望的许多有用的性质可以扩张到条件期望．首先，这里有一些注意事项，条件期望的定义仅仅是对几乎必然相等．回顾$\mathcal{L}^1$的元素是一个可积函数，而 L^1 中的元素是这样函数的等价类．条件期望是 L^1 中的等价类，而不是$\mathcal{L}^1$中的具体函数．通常，这样一个等价类可由它(例如，随机变量 X)中的任意元素表示．条件期望中的等式或不等式都被理解为仅是几乎必然成立．

337

10.1.2 定理 对任意 X，$V\in L^1$，和实数 c，$E(cX+V\mid \mathcal{A})=cE(X\mid \mathcal{A})+E(V\mid \mathcal{A})$.

10.1.3 定理 如果$\mathcal{C}$是$\mathcal{A}$的一个子 σ-代数，那么(由定义可直接得出)对任意 $X\in L^1$，

$$E(X\mid \mathcal{C}) = E(E(X\mid \mathcal{A})\mid \mathcal{C}).$$

对平凡 σ-代数$\{\varnothing,\Omega\}$我们有下面的结论.

10.1.4 命题 对任意的 $X\in L^1$，$E(X\mid \{\varnothing,\Omega\})=EX$.

另一方面，对$\mathcal{A}=\mathcal{S}$，显然有下面的命题成立.

10.1.5 命题 对任意 $X\in L^1(\Omega,\mathcal{S},P)$，$E(X\mid \mathcal{S})=X$.

10.1.6 命题 如果对 L^1 中的函数 f 和 g，$f\leqslant g$ a. s.，那么有 $E(f\mid \mathcal{A})\leqslant E(g\mid \mathcal{A})$.

证明 令 B 表示事件$\{E(f\mid \mathcal{A})>E(g\mid \mathcal{A})\}$．因为条件期望是$\mathcal{A}$-可测函数，所以 B 必定在$\mathcal{A}$中．如果 B 有正的概率，在 B 上对条件期望及函数求积分可得 $E(1_Bf)>E(1_Bg)$，与条件给出的不等式矛盾． □

积分的两个重要的收敛定理对条件期望仍然成立．

10.1.7 条件期望的单调收敛定理 设 f_n，$f\in\mathcal{L}^1$，且 $f_n\uparrow f$ a. s.，那么 $E(f_n\mid \mathcal{A})\uparrow E(f\mid \mathcal{A})$ a. s..

证明 由命题 10.1.6 知，对某个 $g\leqslant E(f\mid \mathcal{A})$，$E(f_n\mid \mathcal{A})\uparrow g$ a. s.．对每个 $A\in\mathcal{A}$，由一般的单调收敛定理(定理 4.3.2)可得，$\int_A g\mathrm{d}P=\int_A f\mathrm{d}P$．因此由唯一性定理 10.1.1 可得，

$$g = E(f\mid \mathcal{A}).$$
□

10.1.8 条件期望的控制收敛定理 如果 $|f_n|\leqslant h\in\mathcal{L}^1$，$f_n\in\mathcal{L}^1$，且 $f_n\to f$ a. s.，那么 $E(f_n\mid \mathcal{A})\to E(f\mid \mathcal{A})$ a. s..

证明 仿照积分控制收敛定理(定理 4.3.5)的证明，令

$$g_n := \inf_{m\geqslant n} f_m \leqslant f_n \leqslant \sup_{m\geqslant n} f_m := h_n.$$

338

那么 $-h\leqslant g_n\uparrow f\leqslant h$，因此根据 10.1.7 可得，$E(g_n\mid \mathcal{A})\uparrow E(f\mid \mathcal{A})$ a. s.．同理，$h\geqslant h_n\downarrow f\geqslant -h$，因此 $E(h_n\mid \mathcal{A})\downarrow E(f\mid \mathcal{A})$ a. s.．因此 $E(g_n\mid \mathcal{A})\leqslant E(f_n\mid \mathcal{A})\leqslant E(h_n\mid \mathcal{A})$ 可以推出 $E(f_n\mid \mathcal{A})\to E(f\mid \mathcal{A})$ a. s.. □

显然，常数的乘积与积分符号可交换．更有趣的是，适当的非常量函数的乘积与条件期望可交换.

10.1.9 定理 如果 g 和 $fg\in\mathcal{L}^1$且 f 关于$\mathcal{A}$是可测的，那么 $E(fg\mid \mathcal{A})=fE(g\mid \mathcal{A})$.

证明 根据定理 10.1.2，假设 $g\geqslant 0$，$f\geqslant 0$. 那么由命题 10.1.6，它们的条件期望是非负的．由命题 4.1.5 知，存在非负简单函数 $f_n\uparrow f$，那么 $f_ng\uparrow fg$，且由定理 10.1.7 可得，$E(f_n\mid \mathcal{A})\uparrow E(f\mid \mathcal{A})$，因此 $gE(f_n\mid \mathcal{A})\uparrow gE(f\mid \mathcal{A})$且 $E(f_ng\mid \mathcal{A})\uparrow E(fg\mid \mathcal{A})$．因此，只要证明 f 是简单函数时结论成立即可．由线性性(定理 10.1.2)，可假设对某个 $C\in\mathcal{A}$，$f=1_C$．因为 $C\cap B\in\mathcal{A}$，那么对任意 $B\in\mathcal{A}$，

$$\int_B 1_C E(g\mid \mathcal{A})\mathrm{d}P = \int_{C\cap B} E(g\mid \mathcal{A})\mathrm{d}P = \int_{C\cap B} g\mathrm{d}P = \int_B 1_C g\mathrm{d}P,$$

因此，由条件期望的唯一性(定理 10.1.1)知，

$$1_C E(g \mid \mathcal{A}) = E(1_C g \mid \mathcal{A}) \text{ a. s.}$$ □

证明 f 和 fg 是可积(如定理 10.1.9 的条件)的一种方法是，只需证明 f 和 g 是可积的且是独立的即可. 另一种方法是证明两者都是平方可积的，对 fg 和 $g \cdot 1$ 应用 Cauchy-Bunyakovsky-Schwarz 不等式(5.1.4)可得 $fg \in \mathcal{L}^1$, $g \in \mathcal{L}^1$.

对任意随机变量 $X \in \mathcal{L}^1$ 和任意子 σ-代数 $\mathcal{A}$, $EX = EE(X \mid \mathcal{A})$，对全空间应用条件期望的定义. 因为在复杂的情况下，计算 EX 可以分成几个部分，首先计算 $E(X \mid \mathcal{A})$，然后再计算它的期望，关系式 $EX = EE(X \mid \mathcal{A})$ 对求期望很有帮助. 这像用累次积分求双重积分(Tonelli-Fubini 定理)一样，但有更广泛的推广. 像选择适当的坐标可以简化积分一样(例如，极坐标或球坐标)，选择适当的 σ-代数 $\mathcal{A}$ 可以简化相应的计算.

最早，条件期望的最重要的情形之一是关于一个随机变量的条件期望. 设 $(\Omega, \mathcal{S}, P)$ 是一个概率空间，X, Y 是 Ω 上的实值随机变量，且 $E|Y| < \infty$. 令 $\mathcal{B}$ 是 $\mathbb{R}$ 中的博雷尔 σ-代数. 那么 $X^{-1}[\mathcal{B}] := \{X^{-1}(A): A \in \mathcal{B}\}$ 是 $\mathcal{S}$ 的子 σ-代数并且是使 X 可测的最小 σ-代数. Y 关于 X 的条件期望记作 $E(Y \mid X)$，定义为 $E(Y \mid X^{-1}[\mathcal{B}])$. 根据定理 4.2.8，对任意两个实值随机变量 X 和 Y，且 $E|Y| < \infty$，存在从 $\mathbb{R}$ 映射到其自身的博雷尔可测函数 g，满足 $E(Y \mid X) = g(X)$，那么 Y 关于 X 的条件期望是 X 的函数.

339

习题

1. 如果 X 是一个随机变量且对于一个常数 c，满足 $|X| \leqslant c$ a. s.，$\mathcal{A}$ 为可测集的任意 σ-代数，证明：$|E(X \mid \mathcal{A})| \leqslant c$ a. s..
2. 设 T 是 $\mathbb{R}^2$ 中的三角形，其中 $0 \leqslant y \leqslant x \leqslant 1$，因此 T 的顶点为 $(0, 0)$, $(1, 0)$ 和 $(1, 1)$，设 P 为 T 上的均匀分布，在 T 上关于平面勒贝格测度的密度为 2，在 T 外为 0. 设 (X, Y) 有分布 P，令 $\mathcal{A}$ 是使 X 可测的最小 σ-代数. 证明：$E(Y \mid \mathcal{A}) = X/2$ a. s..
3. 试举一例：对可测集 A 和 B 及可测集的 σ-代数 $\mathcal{A}$，下面的不等式不成立：
$$E(E(1_A \mid \mathcal{A})E(1_B \mid \mathcal{A})) \leqslant E(1_A 1_B) = P(A \cap B).$$
4. 设 $X_1, X_2, \cdots$ 是独立同分布实值随机变量，满足 $E|X_1| < \infty$, $S_n := X_1 + \cdots + X_n$. 令 $\mathcal{C}_n$ 是使 $S_k (k \geqslant n)$ 可测的最小 σ-代数. 证明：对 $j = 1, \cdots, n$, $E(X_j \mid \mathcal{C}_n) = S_n/n$.
5. 随机变量 X 称为和 σ-代数 $\mathcal{A}$ 独立，当且仅当对 $\mathcal{A}$ 中的每个集合 A 和 X 的值域中的可测集 B, $P(X^{-1}(B) \cap A) = P(X^{-1}(B))P(A)$. 如果 X 是与 $\mathcal{A}$ 独立的实值变量，且 $E|X| < \infty$，证明：$E(X \mid \mathcal{A}) = EX$ a. s..
6. (条件法图引理). 设 $f_n \geqslant 0$ 是概率空间 $(\Omega, \mathcal{S}, P)$ 上的随机变量，对所有的 n, $Ef_n < \infty$，且 $E \liminf_{n\to\infty} f_n < \infty$. 证明：对任意 σ-代数 $\mathcal{A} \subset \mathcal{S}$,
$$E(\liminf f_n / \mathcal{A}) \leqslant \liminf_{n\to\infty} E(f_n / \mathcal{A}).$$
7. 对任意随机变量 $f \geqslant 0$(不必是可积的)和 σ-代数 $\mathcal{A}$，证明：存在随机变量 g，满足 $0 \leqslant g \leqslant +\infty$，关于 $\mathcal{A}$ 是可测的，且对所有的 $A \in \mathcal{A}$, $Ef1_A = Eg1_A$. 证明：可以有 $f < \infty$ a. s.，但是 $g = \infty$ a. s.. [提示：设 Ω 是平面单位元 $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$，设 $\mathcal{A}$ 是使 X 可测的最小 σ-代数.]
8. 设 (X, Y) 是取值于 $\mathbb{R}^2$ 的随机变量. 假设 $\mathcal{L}(X, Y)$ 关于 $\mathbb{R}^2$ 上的勒贝格测度 λ^2 的密度函数为 $f(\cdot,\cdot)$. 设 $\mathcal{C}$ 是使 Y 可测的最小的 σ-代数. 证明：如果 $E|X| < \infty$，
$$E(X \mid \mathcal{C}) = \int_{-\infty}^{\infty} xf(x,Y)\,dx \Big/ \int_{-\infty}^{\infty} f(x,Y)\,dx,$$
其右边是几乎必然有定义的.

340

9. 设$(\Omega, \mathcal{B}, P)$是一概率空间，$(U, \mathcal{G})$，$(V, \mathcal{H})$是两个可测空间，X和Y分别是从Ω映射到U和V的可测函数，f是$U\times V$上的联合可测实值函数且$\int|f(X,Y)|\,\mathrm{d}P(\omega)<\infty$．令$\mathcal{A}$是$\mathcal{B}$的子$\sigma$-代数，假定$X$和$\mathcal{A}$是独立的，$Y$关于$\mathcal{A}$是可测的．在$(U, \mathcal{G})$上，令$\mathcal{L}(X)=\mu$. 证明：

$$E(f(X,Y)\mid\mathcal{A})=\int_U f(x,Y)\,\mathrm{d}\mu(x).$$

10.2 正则条件概率和詹森不等式

对任意概率空间$(\Omega, \mathcal{A}, P)$和子σ-代数$\mathcal{C}\subset\mathcal{A}$，关于$\mathcal{C}$的条件概率定义为：对每个$B\in\mathcal{A}$和$\omega\in\Omega$，$P(B\mid\mathcal{C})(\omega):=E(1_B\mid\mathcal{C})(\omega)$. 对任意$B\in\mathcal{A}$且$P(B)>0$，对每个$A\in\mathcal{A}$，在$\mathcal{A}$上定义概率测度$P(\cdot\mid B)$为$P(A\mid B):=P(A\cap B)/P(B)$. 由定理10.1.2、10.1.7和10.1.8知，对于任意具体的不相交事件序列，可数可加性对几乎所有的ω成立，因而我们也希望条件概率$P(\cdot\mid\mathcal{C})(\omega)$是定义在$\mathcal{A}$上的一个概率测度是合理的．但是零概率集可能依赖于序列，其上可数可加性不成立，并且这些集合的并可能覆盖Ω. 选择适当的条件概率将产生一个条件概率测度，如果它是存在的，定义如下．

设$P_{\mid\mathcal{C}}$是P到$\mathcal{C}$上的限制．定义在$\mathcal{A}\times\Omega$上的函数$P(\cdot\mid\mathcal{C})(\cdot)$称为正则条件概率(regular conditional probability)当且仅当下面两个条件成立：

(a)它是一条件概率．更确切地说，对每个$B\in\mathcal{A}$，$P(B\mid\mathcal{C})(\cdot)=E(1_B\mid\mathcal{C})(\cdot)P_{\mid\mathcal{C}}$几乎必然成立，其中$P(B\mid\mathcal{C})(\cdot)$是关于$\mathcal{C}$可测的，并且，

(b)对$P_{\mid\mathcal{C}}$-几乎每个$\omega\in\Omega$，$P(\cdot\mid\mathcal{C})(\omega)$是$\mathcal{A}$上的一个概率测度．

注：*正则条件概率不一定总是存在的(见习题6)，但是在大多数情况下是存在的，正如下面所述．*

在最后定义的(b)中，“几乎每个”能被“每个”代替：设$C\in\mathcal{C}$，$P(C)=0$时，(a)和(b)成立，对所有的$\omega\notin C$，$P(\cdot\mid\mathcal{C})(\omega)$是一个概率测度．设$\zeta$是$\Omega$中的一个定点并且重新定义：对所有的$\omega\in C$和$B\in\mathcal{A}$，$P(B\mid\mathcal{C})(\omega):=\delta_\zeta(B):=1_B(\zeta)$，那么对所有的$\omega\in\Omega$，(a)和(b)都成立．

341

假设(X, Y)具有在$\mathbb{R}^2$的博雷尔集上的法则(其中X和Y可能不是独立的)，并且想要定义Y关于X(或关于X在其上是可测的最小的σ-代数)的条件分布．目前，条件概率已定义在$\mathbb{R}^2$中的集合上(或者在其他概率空间Ω上，在其上随机变量X和Y必须有定义)．关于$X=x$，在$\mathbb{R}^2$上有一条以y为变量的直线．通常在$\mathbb{R}$中条件分布定义为Y的分布函数要比在$\mathbb{R}^2$的直线上定义为(x, Y)的分布函数简便得多．

例如，假定X和Y仅取整数值．那么，当$P(X=i)>0$时，Y关于X的条件分布可记为$P(Y=j\mid X=i)$. 条件概率的一种记法是$P_{Y\mid X}(j, i):=P(Y=j\mid X=i)$. 这种情况可用多种方法扩展：$Y$可以在某值域空间$T$上取值而不只是在$\mathbb{R}$上取值；$\Omega$不需要是$T$与任意其他空间的积；$Y$和$X$的取值不必是离散的；我们可将在$X$上取值的条件用在$\sigma$-代数$\mathcal{C}$的条件来代替．如果$\mathcal{C}$是$X$在其上可测的最小$\sigma$-代数，$X$是实值变量，那么由定理4.2.8可得，一个函数是$\mathcal{C}$-可测的当且仅当它是$X$的博雷尔可测函数．

这里是随机变量Y在其值域空间T上关于σ-代数$\mathcal{C}$的条件分布的一般定义．令$(\Omega, \mathcal{A}, P)$是一个概率空间，$(T, \mathcal{B})$是一个可测空间．设Y是从Ω映射到T的可测函数．$\mathcal{C}$是$\mathcal{A}$的子σ-代数．那么Y关于$\mathcal{C}$的条件分布是一个从$\mathcal{B}\times\Omega$映射到$[0, 1]$的函数$P_{Y\mid\mathcal{C}}$，满足

(i)对 $P_{\mid \mathcal{C}}$-几乎所有的 ω，$P_{Y\mid \mathcal{C}}(\cdot, \omega)$是$\mathcal{B}$上的概率测度.

(ii)对每个 $B\in\mathcal{B}$，对 $P_{\mid \mathcal{C}}$，$P_{Y\mid \mathcal{C}}(B, \cdot):=P(Y^{-1}(B)\mid \mathcal{C})(\cdot)$几乎必然成立，其中 $P_{Y\mid \mathcal{C}}(B, \cdot)$是$\mathcal{C}$-可测的.

如果$(T, \mathcal{B})=(\Omega, \mathcal{A})$且 Y 是恒等函数 $Y(\omega)\equiv\omega$，那么条件分布 $P_{Y\mid \mathcal{C}}(\cdot,\cdot)$显然给出了一个正则条件概率 $P(\cdot\mid \mathcal{C})(\cdot)$.

$P_{Y\mid \mathcal{C}}$的定义的积空间的情况表明，设$(S, \mathcal{D})$和$(T, \mathcal{B})$是两个可测空间，P 是积空间 $\Omega=S\times T$ 上的概率测度，且积 σ-代数 $\mathcal{A}=\mathcal{D}\otimes\mathcal{B}$. 设 X 和 Y 是从 $S\times T$ 分别映上到 S 和 T 的投影，$X(x, y)\equiv x$，$Y(x, y)\equiv y$. 设$\mathcal{C}=X^{-1}[\mathcal{D}]:=\{X^{-1}(D): D\in\mathcal{D}\}$为 X 在其上是可测的最小 σ-代数. 注意到，对任意 $D\subset S$，$X^{-1}(D)=D\times T$. 在$\mathcal{D}$上，设$\mu:=\mathcal{L}(X):=P\circ X^{-1}$，其称为 P 在 S 上的边缘分布(marginal distribution). 如果下面的条件成立，则称对任意 P 和 $x\in S$，条件分布 P_x 是存在的：

342

(a)对每个 $x\in S$，P_x 是$(T, \mathcal{B})$上的概率测度.

(b)对每个 $B\in\mathcal{B}$，$x\mapsto P_x(B)$是从 S 映射到$\mathbb{R}$ 的$\mathcal{D}$-可测的.

(c)对每个 $D\in\mathcal{D}$，$B\in\mathcal{B}$, $P(D\times B)=\int_D P_x(B)\mathrm{d}\mu(x)$.

如果 P 是积测度 $\mu\times\nu$，那么恰好可取 $P_x\equiv\nu$.

例(典型的条件分布)：

(I)(离散分布) 设 S 和 T 是可数的，$\mathcal{D}=2^S$，$\mathcal{B}=2^T$. 对任意满足 $\mu(\{x\})>0$ 的 $x\in S$，$D\subset T$. 令 $P_x(D):=P(\{\langle x, y\rangle: y\in D\})/\mu(\{x\})$. 固定任意 $t\in T$，对 $\mu(\{x\})=0$，设 $P_x:=\delta_t$，回顾 $\delta_t(B):=1_B(t)$. 这些 P_x 对于 P 是条件分布.

(II)(连续分布) 设 $S=T=\mathbb{R}$，且$\mathcal{D}=\mathcal{B}$是博雷尔 σ-代数. 假设 P 关于$\mathbb{R}^2$ 上的勒贝格测度密度为f. 那么 μ 的密度为 f_X，即关于$\mathbb{R}$ 上的勒贝格测度 λ 有 $f_X(x):=\int_{\infty}^{\infty}f(x,y)\mathrm{d}y$. 对每个 $x\in\mathbb{R}$，$0<f_X(x)<\infty$，$y\in\mathbb{R}$，令 $f_{Y\mid X}(y\mid x):=f(x, y)/f_X(x)$. 设 P_x 是关于 λ 密度函数为 $f_{Y\mid X}(\cdot\mid x)$的法则，如果 $f_X(x)=0$ 或∞，令 P_x 是关于 λ 有固定密度的法则，例如 $1_{[0,1]}$. 那么 P_x 对于 P 是条件分布.

更具体地，设 P 是$\mathbb{R}^2$ 中三角形 V：$0\leqslant y\leqslant x\leqslant 1$ 上的均匀分布，且 $f(x, y)\equiv 21_V(x, y)$. 那么，当 $0\leqslant x\leqslant 1$ 时，$f_X(x)=2x$，否则为 $f_X(x)=0$，并且对 $0\leqslant y\leqslant x$，$0<x\leqslant 1$，$f_{Y\mid X}(y\mid x)=1/x$. P_0 可以任意选取.

条件概率是正则的，或条件分布是存在的，相应的事实是对每个 x，P_x 是可数可加测度——在最后的例子中，对 $0<x\leqslant 1$，$[0, x]$上的均匀分布，即$[0, x]$上的勒贝格测度被 x 整除.

条件分布对把积分(期望)表示为累次积分很有用，甚至可以没有独立性.

10.2.1 定理 (I)如果 P 是积空间上的概率测度，那么条件分布 $P_{Y\mid \mathcal{C}}$存在当且仅当对 P 和 $x\in S$，条件分布 P_x 存在，且对所有的 $B\in\mathcal{B}$，$x\in S$，$y\in T$，

343

$$P_{Y\mid \mathcal{C}}(B,(x,y))=P_x(B)$$

(II)如果这样的条件分布 P_x 存在，那么对任意关于 P 可积的函数 g，

$$\int g\mathrm{d}P=\iint g(x,y)\mathrm{d}P_x(y)\mathrm{d}\mu(x).$$

证明 为证明(I)，由定理 4.2.8 可以得出，$S\times T$ 上的实值函数是$\mathcal{C}$可测的等价于 x 的函数是$\mathcal{D}$-可测的，因而如果条件分布 P_x 存在，则可得 $P_{Y\mid \mathcal{C}}$，其中对所有的 $\omega=(x, y)$，(i)成立. 反之，

如果$P_{Y|\mathcal{C}}$存在，那么由(i)知，存在C且$P(C)=0$，$C\in\mathcal{C}$，因此对某个满足$\mu(D)=0$的$D\in\mathcal{D}$，$C=X^{-1}(D)$，使得对所有的$x\notin D$，$P_{Y|\mathcal{C}}(\cdot(x,y))$是$\mathcal{B}$上的概率测度．固定一个$t\in T$，对所有的$x\in D$，重新定义$P_{Y|\mathcal{C}}(B,(x,y)):=\delta_t(B):=1_B(t)$．那么$P_{Y|\mathcal{C}}$仍是$Y$关于$\mathcal{C}$的条件分布，且现在对所有的$\omega=(x,y)$，(i)成立．由(ii)，对每个$B\in\mathcal{B}$和所有的$(x,y)\in S\times T$，对某个$\mathcal{D}$-可测的$f_B$，$P_{Y|\mathcal{C}}(B,(x,y))=f_B(x)$．令$P_x(B):=f_B(x)$，那么对$P_x$，(a)和(b)成立．下证(c)，对任意$x\in S$，$u\in T$，$B\in\mathcal{B}$，有

$$\begin{aligned}P_x(B)&=P_{Y|\mathcal{C}}(B,(x,u)) &&\text{由 } P_x \text{ 的定义}\\&=P(Y^{-1}(B)\mid\mathcal{C})(x,u) &&\text{a. s.，由(ii) 对 } P_{Y|\mathcal{C}}\\&=E(1_{S\times B}\mid\mathcal{C})(x,u).\end{aligned}$$

对任意$D\in\mathcal{D}$，由μ的定义和对映射X用像测度定理4.1.11，在$D\times T\in\mathcal{C}$上积分最后一个函数得，

$$P(D\times B)=\int_{D\times T}P_x(B)\mathrm{d}P(x,u)=\int_D P_x(B)\mathrm{d}\mu(x).$$

因此得到(c)和(I)成立．

为证(II)，对每个$D\in\mathcal{D}$，$B\in\mathcal{B}$，有$g=1_{D\times B}$．我们可将矩形A转化为有限个不相交矩形的并，然后如4.4节，由单调类(定理4.4.2)，这可推广到一般可测集$A\in\mathcal{A}$，那么$g=1_A$可以由简单函数、非负可测函数和一般可积函数代替．□

拓扑空间$(S,\mathcal{T})$称为波兰空间(Polish space)，当且仅当$\mathcal{T}$对某度量d是可度量化的，使得(S,d)是完全可分度量空间．下面是一个关于正则条件概率的存在性和唯一性的重要定理． [344]

10.2.2 定理 *设T是任意波兰空间，$\mathcal{B}$是其博雷尔集的σ-代数，$(\Omega,\mathcal{A},P)$是任意概率空间，Y是从Ω映射到T的任意可测函数，$\mathcal{C}$是$\mathcal{A}$的任意子σ-代数．那么$\mathcal{B}\times\Omega$上的条件分布$P_{Y|\mathcal{C}}(\cdot,\cdot)$存在．如果$p'(\cdot,\cdot)$也满足$P_{Y|\mathcal{C}}$的定义，那么对$P_{|\mathcal{C}}$-几乎所有的$\omega$，$P'(\cdot,\omega)$和$P_{Y|\mathcal{C}}(\cdot,\omega)$是同一的．在这个意义上，条件分布是唯一的．*

证明 这里$\mathcal{B}$是由可数子集$\mathcal{U}$生成的，例如，存在有理半径且中心为可数稠密子集元素的一切开球(命题2.1.4)．由有限集生成的代数是有限的．因而由$\mathcal{U}$生成的代数$\mathcal{V}$可以写成有限代数集列$\mathcal{V}_n$的可数递增并．由Ulam定理(7.1.4)，在$(T,\mathcal{B})$上将其应用到法则$Q:=P\circ Y^{-1}$上，对每个$B\in\mathcal{V}$，存在紧集$B_j\subset B$，使得当$j\uparrow\infty$时，$Q(B_j)\uparrow Q(B)$．由于紧集的有限并是紧的，对B可选择一个特殊序列$\{B_j\}$，使得$B_1\subset B_2\subset\cdots$，那么由条件的单调收敛性(定理10.1.7)可得，

10.2.3 $$P(Y^{-1}(B_j)\mid\mathcal{C})\uparrow P(Y^{-1}(B)\mid\mathcal{C})\ \text{a. s.}$$

$\mathcal{V}$和对所有$B\in\mathcal{V}$的所有集B_j的并是可数的，因此它生成一个可数代数$\mathcal{D}$．对所有的$D\in\mathcal{D}$，$\omega\in\Omega$，考虑给定的$P(Y^{-1}(D)\mid\mathcal{C})(\omega)$，对每个$D$关于$\omega$是$\mathcal{C}$-可测的，有

(i)对每个$D\in\mathcal{D}$，$P(Y^{-1}(D)\mid\mathcal{C})(\omega)\geqslant 0$a. s.．

(ii)$P(Y^{-1}(T)\mid\mathcal{C})(\omega)=1$a. s.，$P(\varnothing\mid\mathcal{C})(\omega)=0$a. s.．

(iii)对任意$k=1,2,\cdots$和$\mathcal{D}$中不相交的$D_1,\cdots,D_k$，

$$P\Big(Y^{-1}\big(\bigcup_{1\leqslant j\leqslant k}D_j\big)\Big|\mathcal{C}\Big)=\sum_{1\leqslant j\leqslant k}P(Y^{-1}(D_j)\mid\mathcal{C})\quad\text{a. s.}.$$

(iv)对任意$B\in\mathcal{V}$和选择的特殊序列B_j，(10.2.3)几乎必然成立．

现在(i)～(iv)有许多由$\mathcal{C}$-可测函数组成的等式或不等式，也包含了$\mathcal{C}$-可测函数序列的极限，每个都是几乎必然成立的，因而存在某个$W\in\mathcal{C}$，且$P(W)=0$，使得在所有情况下，(i)～(iv)中的

"a. s."可由"对所有的 $\omega \notin W$"来代替.

现在考虑任意两个代数$\mathcal{V} \subset \mathcal{D}$是 S 的子集，其中$(S, \mathcal{T})$是一个豪斯多夫拓扑空间，在$\mathcal{D}$上定义有界、非负、有限可加函数μ，μ在$\mathcal{V}$上关于$\mathcal{D}$是正则的，当且仅当对每个 $B \in \mathcal{V}$,

345

$$\mu(B) = \sup\{\mu(K): K \subset B, K \in \mathcal{D}, K \text{是紧的}\}.$$

如果 $\omega \notin W$，对给定的$\mathcal{V}$和$\mathcal{D}$，$P(Y^{-1}(\cdot) \mid \mathcal{C})(\omega)$成立.

例：在$\mathbb{R}$中，令$\mathcal{V}$是由所有左闭右开区间$[a, b)$生成的代数，令$\mathcal{D}$为所有区间生成的代数．那么，$\mathcal{D}$上任意可数可加的μ在$\mathcal{V}$上关于$\mathcal{D}$是正则的，区间$[a, b)$是由闭区间列$[a, b_n]$$(b_n \uparrow b)$逼近得到.

下面的事实给出了证明可数可加性的一种方法.

10.2.4 引理 如果$\mathcal{V} \subset \mathcal{D}$是两个代数，$\mu$是有限的，且在$\mathcal{D}$上是有限可加的，$\mu$在$\mathcal{V}$上关于$\mathcal{D}$是正则的，那么$\mu$在$\mathcal{V}$上是可数可加的.

证明 反证，假设μ在$\mathcal{V}$上不是可数可加的，那么由定理 3.1.1 知，对某个 $\delta > 0$，有集合 $C_i \in \mathcal{V}$且对所有的 i，$C_i \downarrow \varnothing$，$\mu(C_i) > \delta$. 对每个 i，取一个紧集 $K_i \in \mathcal{D}$，满足 $K_i \subset C_i$，并且 $\mu(C_i \setminus K_i) < \delta/3^i$，那么对每个 n,

$$\mu(K_1 \cap \cdots \cap K_n) \geqslant \mu(C_n) - \sum_{1 \leqslant i \leqslant n} \delta/3^i > \delta/2.$$

因此，每个 $J_n := K_1 \cap \cdots \cap K_n$ 是非空的．如果所有 K_i 的交是空的，那么集合 J_n 的补集构成了 K_1 的一个开覆盖，因此，它们有一个有限子覆盖，但是因为这个补集随着 n 的增加而增大，这意味着某个 J_n 与 K_1 不相交，这与 $J_n \neq \varnothing$矛盾．因此所有 J_n 的交集是非空的，但是与 $C_i \downarrow \varnothing$矛盾． □

2.2 节的习题 17 给出了为什么在最后的证明中豪斯多夫性质是必需的(T_1 是不充分的).

现在继续证明定理10.2.2，对 $\omega \notin W$，$P(Y^{-1}(\cdot) \mid \mathcal{C})(\omega)$在代数$\mathcal{V}$上是可数可加的，并将其扩张为$\sigma$-代数$\mathcal{B}$上的可数可加概率测度$\mu_\omega$(定理 3.1.4). 对 $\omega \in W$，设μ_ω为任意固定的概率测度，例如 Q. 令$\mathcal{E}$是所有集合 $B \in \mathcal{B}$的并，使得 $\omega \mapsto \mu_\omega(B)$满足 $P(Y^{-1}(B) \mid \mathcal{C})$的定义．那么$\mathcal{E}$包含了代数$\mathcal{V}$，且$\mathcal{E}$是单调类(定理 10.1.7)，由定理 4.4.2 知，$\mathcal{E}$包含了所有的σ-代数$\mathcal{B}$. 对 $B \in \mathcal{B}$和 $\omega \in \Omega$，$P_{Y \mid \mathcal{C}}(B, \omega) := \mu_\omega(B)$定义了一个条件分布. $W \in \mathcal{C}$可得到在 ω 的$\mathcal{C}$-可测性．定理 10.2.2 的存在性证毕.

下证唯一性，由 $P_{\mid \mathcal{C}}$-几乎必然性保证了对生成$\mathcal{B}$的代数$\mathcal{V}$的所有集合，两个条件分布是相等的．对所有这样的 ω，在这些集合的集族上存在一个相同的单调类，因此它包括了所有的$\mathcal{B}$，定理 10.2.2 得证． □

346

函数 h 关于测度μ的积分通常记为$\int h \mathrm{d}\mu$或$\int h(x)\,\mathrm{d}\mu(x)$，有时记为$\int h(x)\mu(\mathrm{d}x)$．这样能明显地看出积分变元，当是条件分布时，测度$\mu$依赖于其他参数．下面将给出当条件分布存在时，对每个 ω，条件期望可写成积分形式.

10.2.5 定理 设$(\Omega, \mathcal{A}, P)$是一个概率空间，$(T, \mathcal{B})$是一个可测空间，Y是从 Ω 映射到 T 的可测函数，$\mathcal{C}$是$\mathcal{A}$的一个子σ-代数，$P_{Y \mid \mathcal{C}}(\cdot, \cdot)$是$\mathcal{B} \times \Omega$上的条件分布，$g$为从 T 映射到$\mathbb{R}^k$的可测函数，满足 $E \mid g(Y) \mid < \infty$，那么对 $P_{\mid \mathcal{C}}$-几乎所有 ω，g 关于 $P_{Y \mid \mathcal{C}}(\cdot, \omega)$是可积的，且

$$E(g \circ Y \mid \mathcal{C})(\omega) = \int g(y) P_{Y \mid \mathcal{C}}(\mathrm{d}y, \omega).$$

证明 如果定理对 g 的每一分量成立，那么它对 g 也是成立的，因此可假设 $k = 1$. 首先假设对

某个 $B\in\mathcal{B}$，$g=1_B$，那么

$$E(1_B\circ Y\mid \mathcal{C})(\omega)=E(1_{Y^{-1}(B)}\mid \mathcal{C})(\omega)\text{a. s.}$$

且由定义可得，

$$\int 1_B(y)P_{Y|\mathcal{C}}(\mathrm{d}y,\omega)=P_{Y|\mathcal{C}}(B,\omega)=P(Y^{-1}(B)\mid \mathcal{C})(\omega).$$

因为定理10.2.5中的方程两边是 g 的线性函数，故对任意简单函数 g 结论均成立．如果 $g\geqslant 0$，令 $0\leqslant g_n\uparrow g$，g_n 为简单函数．那么 $0\leqslant g_n\circ Y\uparrow g\circ Y$. 由假设得，$E\mid g(Y)\mid$ 是有限的，因此由条件期望的单调收敛性(定理10.1.7)可得，$E(g_n\circ Y\mid \mathcal{C})\uparrow E(g\circ Y\mid \mathcal{C})$a. s.，并且它的取值是有限的．当 g_n 收敛到 g 时，条件分布的单调收敛性定理对每个 ω 给出定理右边的积分的收敛性．因此对于可测的 $g\geqslant 0$，定理成立．由线性性和 $g=g^+-g^-$ 知，定理对所有满足 $E\mid g\circ Y\mid<\infty$ 的 g 都成立． □

对于下面的结论，回顾(6.2~6.3节)一个集合 $C\subset\mathbb{R}^k$ 称为凸的(convex)，当且仅当对每个 x，$y\in C$，且 $1<p<1$，有 $px+(1-p)y\in C$. C 上取值于$[-\infty,\infty]$的函数 f 称为凸的，当且仅当对所有的 x，y，p，

$$f(px+(1-p)y)\leqslant pf(x)+(1-p)f(y).$$

347

如果 f 在凸集 C 上是凸的，其闭包为 $\bar{C}$，在 $\bar{C}$ 上定义 $\bar{f}$，在 C 上 $\bar{f}=f$，在 $\bar{C}\setminus C$ 上 $\bar{f}=+\infty$，那么很容易得出 $\bar{f}$ 在 $\bar{C}$ 上是凸的．

当 f 是线性函数且 EX 有限时，$Ef(X)=f(EX)$. 对非线性函数，仅当函数具有适当的性质时，我们可得出一个不等式．例如，如果 X 是实值的且 $E\mid X\mid<\infty$，将 Cauchy-Bunyakovsky-Schwarz 不等式(5.1.4) 应用到 $X\cdot 1$ 可得，$(EX)^2\leqslant E(X^2)$. 因此对 $f(x)=x^2$，有 $f(EX)\leqslant Ef(X)$. 这里 f 是一个凸函数，并且这个不等式扩展到一般的凸函数上有下面的结论.

10.2.6 詹森不等式 设 C 是 $\mathbb{R}^k$ 中非空的博雷尔可测凸集，f 是定义在 C 上的凸函数，X 是在 C 上取值的随机变量，满足 $E\mid X\mid<\infty$，且 $f(X)$ 是随机变量．那么 $EX\in C$，$Ef(X)$ 有定义，且 $-\infty<Ef(X)\leqslant+\infty$，$Ef(X)\geqslant f(EX)$.

注：如果 C 是开集，或 X 在 C 的内部 U 中取值，其中 U 是凸的，那么因为 f 是 U 上的连续函数(定理6.3.4)，从其他的假设条件可得 $f(X)$ 是可测的．另一方面，例如，设 U 是开圆盘 $\{\langle x,y\rangle: x^2+y^2<1\}$，设 C 是 U 和 T 的并，其中 T 为圆 $x^2+y^2=1$ 的任意子集．设 f 是定义在 U 上的有界凸函数，例如 $f\equiv 0$ 或 $f(x,y)=(x^2+y^2)^{1/2}$，因此在 U 上 $f\leqslant 1$. 设 f 是定义在 T 上的任意函数且对所有的 $z\in T$，$f(z)\geqslant 1$. 那么 f 是凸的但不必是连续的，甚至不必是博雷尔可测的，因此即使 T 是全部的单位圆，如果 X 在 T 中取值的概率是正的，$f(X)$ 可以不是可测的．

证明 易见闭包 $\bar{C}$ 是凸的，因此，它是包含它的所有闭半空间的交集(定理6.2.9). 这些半空间具有形式 $\{x: h(x)\geqslant t\}$，其中 h 是 $\mathbb{R}^k$ 上的非零线性函数．那么 $Eh(X)=h(EX)$. 显然 EX 在每个包含 $\bar{C}$ 的半空间上，因此 $EX\in\bar{C}$. 如果 EX 在 C 的边界上，那么在 EX 上对 C 选择一个支撑超平面(定理6.2.7)，我们有一个线性的 $h\neq 0$，$t\in\mathbb{R}$，对所有的 $x\in C$，$h(x)\geqslant t$，所以 $h(X)\geqslant t$，且 $Eh(X)=h(EX)=t$. 那么 $h(X)=t$a. s.. 我们可用超平面 $h^{-1}\{t\}$ 代替 $\mathbb{R}^k$，用 $C\cap h^{-1}\{t\}$ 代替 C，从 k 维降到1维．继续用这种方法，如果 $k=1$，那么 C 是一个区间，如果 EX 在其边界上，那么 X 几乎必然是一个常数，且定理成立．因此可假设 EX 在 C 的内部 U 上，那么它是非空的与命题6.2.10中闭包 $\bar{C}$ 的内部一样．

348

令 $D:=\{\langle x, y\rangle\in\mathbb{R}^{k+1}: x\in C, y\geqslant f(x)\}$，由凸函数的定义(6.3.1)知，$D$ 是一个凸集．显然，在 D 的边界上，$V:=\langle EX, f(EX)\rangle$．由定理 6.2.7 知，$D$ 在 V 上有一个支撑超平面 H，如 $a_0y+g(x)=c$，其中 g 是$\mathbb{R}^k$ 上的线性函数，$g(x)\equiv\sum_{1\leqslant j\leqslant k}c_jx_j$．如果 $a_0=0$，那么 $g\neq 0$．因为 EX 是 U 中的开集，那么在 C 中存在使得 $g(x)>g(EX)=c$ 的点 x，以及另一些使得 $g(x)<c$ 的点 x，取 y 足够大，在 D 上存在使得 $g(x)>c$ 的点$\langle x, y\rangle$，以及使得 $g(x)<c$ 的另一些点，这与 H 是 D 的支撑超平面矛盾(例如，如果 $k=1$，直线 $g(x)=c$ 是垂直的并且把 D 中的点分为两部分)．因此 $a_0\neq 0$．由 a_0 等分，可假设 $a_0=1$．那么 D 被包含在闭半空间$\{\langle x, y\rangle: y\geqslant c-g(x)\}$内．对所有的 $x\in C$，有 $f(x)\geqslant c-g(x)$，其中 $f(EX)=c-g(EX)$．因而

$$f(X)-f(EX)\geqslant g(EX)-g(X).$$

由于 g 是线性的，右边是可积的且积分为 0．由假设可得 $f(X)$是可测的．因此，$f^-(X)$有有限积分并且 $Ef(X)$有定义，可能为 $+\infty$．式子两边取期望得 $Ef(X)\geqslant f(EX)$． □

10.2.7 条件詹森不等式 设$(\Omega, \mathcal{A}, P)$是一个概率空间，且 f 是 Ω 上取值于$\mathbb{R}^k$ 中开凸集 C 的随机变量．设 g 为定义在 C 上的实值凸函数．如果 $|f|$ 和 $g\circ f$ 是可积的，且$\mathcal{C}$是$\mathcal{A}$的任意子 σ-代数，那么 $E(f|\mathcal{C})\in C$a. s.，且

$$E(g(f)|\mathcal{C})\geqslant g(E(f|\mathcal{C})).$$

证明 当上述 C 是完备可分度量空间中的开集时，它是波兰空间(定理 2.5.4)．因此，存在条件分布 $P_{f|\mathcal{C}}(\cdot,\cdot)$(定理 10.2.2)．我们可以将条件期望写成关于条件分布的积分(定理 10.2.5)：

$$E(f|\mathcal{C})(x)=\int_C yP_{f|\mathcal{C}}(\mathrm{d}y,x)\text{ a. s.}$$

$$E(g\circ f|\mathcal{C})(x)=\int_C g(y)P_{f|\mathcal{C}}(\mathrm{d}y,x)\text{ a. s.}$$

那么对几乎所有的 x 将非条件詹森不等式(10.2.6)运用到法则 $P_{f|\mathcal{C}}(\cdot, x)$可得，$g(E(f|\mathcal{C})(x))\leqslant E(g\circ f|\mathcal{C})(x)$． □

[349] 如果 T 是一个波兰空间，应用定理 10.2.2 和定理 10.2.1 我们很容易得到下面的结论．

10.2.8 命题 设$(S, \mathcal{D})$是一个可测空间且 T 是具有博雷尔 σ-代数$\mathcal{B}$的波兰空间．设从 $\omega:=S\times T$ 映射到 T 为 $Y(x, y):=y$．设$\mathcal{A}$是 Ω 上的积 σ-代数，$\mathcal{C}$是子 σ-代数$\{D\times T: D\in\mathcal{D}\}$．设 P 是$\mathcal{A}$上任意概率测度．那么存在条件分布 $P_{Y|\mathcal{C}}(\cdot,\cdot)$，其中 $P_x:=P_{Y|\mathcal{C}}(\cdot, (x, y))$不依赖于 y．换句话说，条件分布 P_x 存在．

证明 满足定理 10.2.2 的条件，因此，在$\mathcal{B}\times\Omega$上存在条件分布 $P_{Y|\mathcal{C}}$．对每个 $B\in\mathcal{B}$，$(x, y)\mapsto P_{Y|\mathcal{C}}(B, (x, y))$是$\mathcal{C}$-可测的，因此它不赖于 y． □

下面是关于条件期望的性质的一个例子：给定希尔伯特空间 H 的一个闭线性子空间 F，对 H 中的每个点 z，F 中存在唯一的 x，使得 $z-x$ 与 F 是正交的(定理 5.3.8)．从 z 映射到 x 的函数显然是线性的，称为从 H 映上到 F 的正交投影．对平方可积函数，条件期望是这样一个正交投影．

10.2.9 定理 设$(\Omega, \mathcal{B}, P)$是任意概率空间，$\mathcal{C}$为$\mathcal{B}$的任意子 σ-代数．那么 $F:=L^2(S, \mathcal{C}, P)$是希尔伯特空间 $H=L^2(S, \mathcal{B}, P)$的一个闭线性子空间，在 H 上关于$\mathcal{C}$的条件期望是从 H 映上到 F 的正交投影．

证明 由定理 5.2.1 和定理 5.3.1 知，L^2 空间是希尔伯特空间，F 是 H 的一个闭线性子空间．对任意 $X\in H$，对 X 和 1 运用 Cauchy-Bunyakovsky-Schwarz 不等式得 $X\in\mathcal{L}^1$．那么对 $g(t)=t^2$ 运用条

件詹森不等式(10.2.7)可得，$Y := E(X \mid \mathcal{C}) \in \mathcal{L}^2$. 因此 $Y \in F$. 对任意 $V \in F$，运用 Cauchy-Bunyakovsky-Schwarz 不等式可得，$E|V(X-Y)|$ 是有限的，并且由定理 10.1.9 得，

$$E(V(X-Y)) = EE(V(X-Y) \mid \mathcal{C}) = E(VE(X-Y \mid \mathcal{C})),$$

因此，$X-Y$ 与 F 是正交的，且如同定理 5.3.8 中那样，$X = Y + (X-Y)$ 是 X 的正交分解. □

例(博雷尔悖论)：设 S^2 在单位球面 $\{(x, y, z): x^2+y^2+z^2=1\}$ 上，在 S^2 上有球坐标 θ, ϕ，满足 $-\pi < \theta \leqslant \pi$, $-\pi/2 \leqslant \phi \leqslant \pi/2$,

$$(x,y,z) = (\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi).$$

设 P 为 S^2 上的均匀概率，等于表面积的 $1/(4\pi)$，其中

$$dP(\theta,\phi) = \cos\phi d\phi d\theta/(4\pi) = (\cos\phi d\phi/2)(d\theta/(2\pi))$$

350

是一个积测度，但是 P 是有理不变量，因此不依赖于坐标的选择. P 关于大圆 K 的条件分布是什么?

(a) θ 关于 ϕ(平行于纬线圈上)的条件分布能全部取为均匀分布 $d\theta/(2\pi)$, $-\pi < \theta \leqslant \pi$. 例如，对 $\phi = 0$(赤道，最大圆).

(b) ϕ 关于 θ(在经线圈上)的条件分布能全部取为 $\cos d\phi/2$，即在半大圆圈上的非均匀分布. 两个这样的半圆 θ 的值与大圆上 π 的形式是不同的.

因而，P 在大圆上 K(集合 $P(K)=0$)上的条件分布不是唯一确定的且依赖于坐标的选择. 在(a)或(b)中，一个坐标的条件分布可由几乎处处相等的另一个坐标唯一确定(定理 10.2.2).

习题

1. 设 P 是 $\mathbb{R}^2$ 上的概率测度，其关于勒贝格测度 λ^2 的密度函数为 $f(x, y)$. 令 $\mathcal{A}$ 是 $\mathbb{R}^2$ 上的博雷尔 σ-代数，$\mathcal{C}$ 是 X 在其上是可测的最小 σ-代数. 如果当 $0 \leqslant y \leqslant x \leqslant 1$ 时，$f(x, y) = 3x^2+3y^2$；其他情况时，$f(x, y) = 0$，试求：y 关于 P 的条件密度.
2. 证明：对 $1 \leqslant p < \infty$，任意 $f \in \mathcal{L}^p(X, \mathcal{S}, P)$ 和 σ-代数 $\mathcal{C} \subset \mathcal{A}$，有
$$|E(f \mid \mathcal{C})| \leqslant E(|f|^p \mid \mathcal{C}).$$
3. (a) 设 $f > 0$ 是使得 $\log f$ 和 f 都在 $\mathcal{L}^1$ 中的函数. 证明：对任意子 σ-代数 $\mathcal{C}$,
$$E(\log f \mid \mathcal{C}) \leqslant \log E(f \mid \mathcal{C}).$$
(b) 试举例：$\mathcal{L}^1$ 中的随机变量 $f > 0$，使得 $\log f$ 不在 $\mathcal{L}^1$ 中. [提示：令 $f \leqslant 1$.]
4. 举例：在开区间(0, 1)上的实值凸函数 f，满足 $\int_0^1 f(x)\,dx = +\infty$. [提示：令 $f(x) = 1/x$.]
5. 设 $(X, \mathcal{A}, P)$ 是一个概率空间，$\mathcal{C}$ 为 $\mathcal{A}$ 中所有使得 $P(A) = 0$ 或 1 的集合 A 的集族. 证明：$\mathcal{C}$ 是一个 σ-代数，并求出正则条件概率 $P(\cdot \mid \mathcal{C})(\cdot)$.
6. 求一个概率空间 $(X, \mathcal{A}, P)$ 和子 σ-代数 $\mathcal{C} \subset \mathcal{A}$，使得在其上不存在正则条件概率 $P(\cdot \mid \mathcal{C})(\cdot)$. [提示：令 $X = [0, 1]$ 具有博雷尔 σ-代数 $\mathcal{C}$，P 为勒贝格测度.] 令 C 内测度为 0，外测度为 1(定理 3.4.4). 令 $\mathcal{A}$ 是由 $\mathcal{C}$ 和 C 生成的，且对所有的 A, $B \in \mathcal{C}$，有 $P((A \cap C) \cup (B \setminus C)) = (P(A) + P(B))/2$. [提示：由定理 10.2.2 中的唯一性知，在 $\mathcal{C}$ 上有 $P(\cdot \mid \mathcal{C})(x) = \delta_x$，因而在 $\mathcal{A}$ 上对 P-几乎所有的 x 成立.]

351

7. 设 f 是开区间 (a, b) 上的实值函数且 f 是非凸的. 证明：对某个随机变量 $X(a < X < b)$, $f(EX) > Ef(X)$.
8. 证明：$\mathbb{R}^k$ 中的任意开凸集 C 对于线性(仿射)函数 g 和 $t \in \mathbb{R}$ 是开半空间 $\{x: g(x) > t\}$ 的交. [提示：参考詹森不等式 10.2.6 的证明. 考虑 $C = \mathbb{R}^k$ 作为空集的交.]
9. 证明：如果 C 是 $\mathbb{R}^1$ 中的任意凸集，f 是 C 上的任意凸函数，X 是在 C 上取值的随机变量，那么 $f(x)$ 总是可测的(一个随机变量).
10. 设 X 和 Y 是两个独立同分布随机变量，$X > 0$ a. s. 且 $EX < \infty$. 证明：$E(Y/X) > 1$，除非对某个常数 c，$X = c$ a. s.. 举例说明 $E(Y/X) = \infty$. (这看起来可能是矛盾的：X 和 Y 是同分布的，在某种意义上大小是相同的，

但是 $E(Y/X)>1$ 将暗示了 Y 比 X 要大，而对 $E(X/Y)>1$ 也是如此).

11. 证明或举反例：对任意集合 X，σ-代数 $\mathcal{A}\subset\mathcal{B}$，$\mathcal{B}$是 X 的子集，P 和 Q 是$\mathcal{B}$上的概率测度，$0<t<1$，可选择 $P(\cdot\mid\mathcal{A})$和 $Q(\cdot\mid\mathcal{A})$，使得对每个 $B\in\mathcal{B}$，$(tP+(1-t)Q)(B\mid\mathcal{A})=tP(B\mid\mathcal{A})+(1-t)Q(B\mid\mathcal{A})$ 对 $P+Q$ 几乎处处成立.
12. 对 $\forall x_j\geq 0$，由詹森不等式证明：算术—几何平均不等式(5.1.6)$(x_1x_2\cdots x_n)^{1/n}\leq(x_1+\cdots+x_n)/n$.［提示：$x_j=0$ 显然成立，如果所有的 $x_j>0$，利用变换 $y_j=\log(x_j)$.］
13. 设 $X:=(X_1,\cdots,X_k)$在$\mathbb{R}^k$ 上服从正态分布 $N(m,C)$. 设$\mathcal{A}$是 $X_2,\cdots,X_k$ 在其上可测的最小 σ-代数. 证明：对某些常数 $c_1,\cdots,c_k$，$E(X_1\mid\mathcal{A})=c_1+\sum_{2\leq j\leq k}c_jX_j$.［提示：首先假定 $m=0$(那么 c_1 也为0)结论成立. 设 U 是由 $X_2,\cdots,X_k$ 生成的随机变量的线性空间. 令 Y 是由 X_1 映射到 U 的正交投影. 那么 $V:=X_1-Y$ 与 $X_2,\cdots,X_k$ 是正交的，利用联合特征函数 $E\exp(\mathrm{i}(t_1,\cdots,t_k)\cdot(V,X_2,\cdots,X_k))$和 9.5 节的结论证明它与$(X_2,\cdots,X_k)$是独立的.］
14. 称集合 C 上的函数f是*严格凸的*(strictly convex)，如果对 C 中任意 $x\neq y$ 和 $0<t<1$，$f(tx+(1-t)y)<tf(x)+(1-t)f(y)$. 证明：在詹森不等式(10.2.6)中，除了 $X=EX$a. s. 外，$Ef(X)>f(EX)$.

352

10.3 鞅

给定一个集合 T、可测空间$(S,\mathcal{S})$和概率空间$(\Omega,\mathcal{B},P)$，*随机过程*(stochastic process)是取值于 S 的函数$\langle t,\omega\rangle\mapsto X_t(\omega)$，其中 $t\in T$，$\omega\in\Omega$，使得对每个 $t\in T$，$X_t(\cdot)$是可测的. 通常 $S=\mathbb{R}$ 且$\mathcal{S}$是博雷尔集的 σ-代数. T 通常是实直线的子集，被当作时刻集，因此，X_t 是某质点在 t 时刻的值.

假定$(T,\leq)$是线性序且$\{\mathcal{B}_t\}_{t\in T}$是 σ-代数族，对 $t\leq u$，$\mathcal{B}_t\subset\mathcal{B}_u\subset\mathcal{B}$. 那么$\{X_t,\mathcal{B}_t\}_{t\in T}$称为*鞅*(martingale)，当且仅当对所有的 t，$E\mid X_t\mid<\infty$，且当 $t\leq u$ 时

10.3.1
$$X_t=E(X_u\mid\mathcal{B}_t).$$

如果在(10.3.1)中的“=”由“≤”或“≥”代替，且对所有的 t，X_t 是$\mathcal{B}_t$可测的，那么称$\{X_t,\mathcal{B}_t\}$为*下鞅*(submartingale)或*上鞅*(supermartingale).

本书中 T 通常是指正整数集 $n=1,2,\cdots$那么$\{X_n,\mathcal{B}_n\}_{n\geq 1}$称为*鞅序列*(martingale sequence，或下鞅序列，或上鞅序列).

如果我们将 X_t 视为一个赌徒在 t 时刻的运气，那么在这种意义上，一个鞅序列表明在任意时刻 t 游戏是公平的. 现在的结果(对于给定的$\mathcal{B}_t$)与以前的结果无关. 如果游戏的时间较长，则赌徒净赚或净赔的期望值 u 为 0. 同理，一个下鞅将赞成玩家继续赌下去，而一个上鞅则反对玩家继续游戏.

例：设 $X_1,X_2,\cdots$是独立实值随机变量，$S_n:=X_1+\cdots+X_n$，$\mathcal{B}_n$是 $X_1,\cdots,X_n$ 在其上可测的最小 σ-代数. 那么对每个 n，X_{n+1}与$\mathcal{B}_n$中的每一个事件都是独立的：设$\mathcal{C}_n$是所有与 X_{n+1}独立的事件的并，那么由独立性的定义知，对$\mathbb{R}$ 中的博雷尔集 A_j 和$j\leq n$，$\mathcal{C}_n$包含了所有事件 $X_j^{-1}(A_j)$及所有这些事件的交. $\mathcal{C}_n$中集合的有限不相交并仍包含在$\mathcal{C}_n$中. 因而由事件 $X_j^{-1}(A_j)$生成的代数$\mathcal{A}$仍包含在$\mathcal{C}_n$中(由命题 3.2.2 和命题 3.2.3 及归纳法). 显然$\mathcal{C}_n$是单调类，因此由定理 4.4.2 知，它包含了由$\mathcal{A}$生成的 σ-代数，那正是$\mathcal{B}_n$所需要的.

因而如果对所有的 n，$EX_n=0$，那么$\{S_n,\mathcal{B}_n\}$是鞅；如果 $EX\geq 0$，则它是下鞅；如果 $EX\leq 0$，则它是上鞅.

迭代条件期望(10.1.3)并归纳给出下面的命题.

10.3.2 命题 序列$\{X_n, \mathcal{B}_n\}$是鞅，当且仅当对所有的n，$\mathcal{B}_n \subset \mathcal{B}_{n+1}$，$E|X_n| < \infty$，且$X_n = E(X_{n+1} | \mathcal{B}_n)$，同理对下鞅或上鞅，$X_n$是$\mathcal{B}_n$可测的，“=”分别由“≤”或“≥”代替即可. [353]

某些适当的函数保持下鞅的性质.

10.3.3 定理 设f是$\mathbb{R}$中开区间U上的凸函数，且$\{X_t, \mathcal{B}_t\}_{t\in T}$是下鞅，使得对所有的$t$，$X_t$取值于$U$且$E|f(X_t)| < +\infty$. 如果或者

(a)f是非减的，或者

(b)$\{X_t, \mathcal{B}_t\}$是鞅，则$\{f(X_t), \mathcal{B}_t\}_{t\in T}$是下鞅.

证明 对$t \le u$，在情况(a)中，因为$X_t \le E(X_u | \mathcal{B}_t)$且$f\uparrow$，$f(X_t) \le f(E(X_u | \mathcal{B}_t))$. 在情况(b)中，$f(X_t) = f(E(X_u | \mathcal{B}_t))$. 在其他情况下，由条件詹森不等式(10.2.7)可得，$f(E(X_u | \mathcal{B}_t)) \le E(f(X_u) | \mathcal{B}_t)$. □

例：如果$\{X_t, \mathcal{B}_t\}_{t\in T}$是一鞅，那么$\{|X_t|, \mathcal{B}_t\}_{t\in T}$是一下鞅.

下面的事实有助于把下鞅或上鞅收敛性的研究归约到鞅.

10.3.4 定理(Doob 分解) 对任意下鞅序列$\{X_n, \mathcal{B}_n\}$，存在随机变量Y_n和Z_n，且对所有的n，$X_n \equiv Y_n + Z_n$，其中$\{Y_n, \mathcal{B}_n\}$是鞅，$Z_1 \equiv 0$，当$n \ge 2$时，Z_n是$\mathcal{B}_{n-1}$可测的，且对所有的n和几乎对所有的ω，$Z_n(\omega) \le Z_{n+1}(\omega)$. 满足这些性质的$Y_n$和$Z_n$是唯一确定的.

注：这样的序列$\{Z_n\}$称为递增过程. 对每个n，$E|Z_n| \le E|X_n| + E|Y_n| < +\infty$. 假设$X_1 \le X_2 \le \cdots \le X_n \le \cdots$其中$X_n$是$\mathcal{B}_n$可测的. 那么$\{X_n\}$是下鞅，且在一定意义上是一个递增过程，但是$\{X_n\}$的唯一 Doob 分解不是$X_n = 0 + X_n$(见证明后的例子)，除非$X_n$是$\mathcal{B}_{n-1}$可测的.

证明 令$D_1 := X_1$，$D_n := X_n - X_{n-1}$，$n \ge 2$. 令$G_1 := 0$，$G_n := E(D_n | \mathcal{B}_{n-1})$，$n \ge 2$. 那么对所有的$n$，$G_n \ge 0$a.s.. 因为$\{X_n\}$是下鞅，令$H_n := D_n - G_n$， [354]

$$Z_n := \sum_{j=1}^{n} G_j, \quad Y_n := X_n - Z_n = \sum_{j=1}^{n} H_j.$$

那么$\{Z_n\}$是一个递增过程. 对$n \ge 2$，

$$E(H_n | \mathcal{B}_{n-1}) = E(D_n | \mathcal{B}_{n-1}) - E(G_n | \mathcal{B}_{n-1}) = G_n - G_n = 0.$$

因而$E(Y_{n+1} | \mathcal{B}_n) = E(Y_n | \mathcal{B}_n) = Y_n$，且$\{Y_n, \mathcal{B}_n\}$是鞅，因此具有给定性质的$Y_n$和$Z_n$的存在性得证.

下证唯一性，设$\{y_n\}$和$\{z_n\}$是另外两个与$\{Y_n\}$和$\{Z_n\}$具有相同性质的序列. 那么$z_1 = Z_1 \equiv 0$，因此$y_1 = Y_1 = X_1$. 为用归纳法证明，假定$z_j = Z_j$，$y_j = Y_j$，$j = 1, \cdots, n-1$. 那么

$$\begin{aligned} z_n &= E(z_n | \mathcal{B}_{n-1}) \quad \text{因为 } z_n \text{ 是}\mathcal{B}_{n-1}\text{可测的} \\ &= E(X_n - y_n | \mathcal{B}_{n-1}) = E(X_n | \mathcal{B}_{n-1}) - y_{n-1} \\ &= E(X_n | \mathcal{B}_{n-1}) - Y_{n-1} \\ &= E(X_n - Y_n | \mathcal{B}_{n-1}) = E(Z_n | \mathcal{B}_{n-1}) \\ &= Z_n \end{aligned}$$

从而$Y_n = y_n$，即唯一性得证. □

例：设$X_1 \equiv 0$，$X_2 = 2$或4的概率各为1/2. 那么$X_1 \le X_2$. 另一方面，令$Z_2 = 3$，$Y_2 = X_2 - Z_2$. 令$\mathcal{B}_1 := \{\varnothing, \Omega\}$是平凡的$\sigma$-代数. 令$Y_1 = 0$. 那么$X_n = Y_n + Z_n (n = 1, 2)$是定理10.3.4给出的分解. 除了$X_2$不是$\mathcal{B}_1$可测的外，$X_n = 0 + X_n$的另一个这样的分解，要证明分解的唯一性，证明$Z_n$的$\mathcal{B}_{n-1}$可测性是必要的.

随机变量的集合$\{X_t\}_{t\in T}$称为L^1有界的(L'-bounded)，当且仅当$\sup_{t\in T}E|X_t|<+\infty$. 如果

$$\lim_{M\to\infty}\sup_{t\in T}E(|X_t|1_{\{|X_t|>M\}})=0,$$

则$\{X_t\}$称为一致可积的.

例：序列$n1_{[0,1/n]}$是L^1有界的但对[0, 1]上的勒贝格测度不是一致可积的.

10.3.5 定理 $\{X_t\}$是一致可积的当且仅当它是L^1有界的，对每个$\varepsilon>0$，存在$\delta>0$，使得对每个事件A，有$P(A)<\delta$，对所有的t，有$E(|X_t|1_A)<\varepsilon$.

355

证明 充分性：对所有的t，令$E|X_t|\leqslant K<\infty$，给定$\varepsilon>0$，取相应的δ. 对于$M>K/\delta$，对所有的t，$P(|X_t|>M)<\delta$，那么

$$E(|X_t|1_{\{|X_t|>M\}})<\varepsilon.$$

必要性：如果$\{X_t\}$是一致可积的，对所有的t，M满足$E|X_t|1_{\{|X_t|>M\}}<\varepsilon$，那么对所有的$t$，$E|X_t|<M+1$，因此$\{X_t\}$是$L^1$有界的. 给定$\varepsilon>0$，如果$K$足够的大，使得对所有的$t$，

$$E(|X_t|1_{\{X_t|>K\}})<\varepsilon/2,$$

取$\delta<\varepsilon/(2K)$. 那么当$P(A)<\delta$时，

$$E|X_t|1_A\leqslant E|X_t|1_A1_{\{X_t|\leqslant K}+E|X_t|1_{\{X_t|>K}<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$ □

如果$f\in\mathcal{L}^1$且对所有的t，$|X_t|\leqslant f$，因为对所有的t，及任意的M，有$\{|X_t|>M\}\subset\{f>M\}$，故容易得出$\{X_t\}$是一致可积的. 但是对于$L^1$中的一个函数，不是所有的一致可积族都是绝对有界的：例如，在具有勒贝格测度λ的[0, 1]上，考虑所有具有$n1_A$形式的函数的集合，其中A是任意满足$\lambda(A)=1/n^2(n\geqslant1)$的集合. 下面是一个改进的控制收敛定理. 给出了一个不能再改进的充要条件.

10.3.6 定理 $\mathcal{L}^1$中的X_n，X，当$n\to\infty$时，$E|X_n-X|\to0$当且仅当X_n依概率收敛到X且$\{X_n\}_{n\geqslant1}$是一致可积的.

证明 充分性：一个子列$X_{n(k)}\to X$a.s.(由定理9.2.1)，因此由法图引理(4.3.3)知，令$X_0:=X$，$\{X_n\}_{n\geqslant0}$是一致可积的. 给定$\varepsilon>0$，取$\delta>0$，使得$P(A)<\delta$蕴涵着对所有的$n\geqslant0$，有$E(|X_n|1_A)<\varepsilon/4$，取足够大的$n_0$，使得对所有的$n\geqslant n_0$，$P(|X_n-X|>\varepsilon/2)<\delta$. 那么

$$E|X_n-X|\leqslant\varepsilon/2+\varepsilon/2=\varepsilon.$$

必要性：如果$E|X_n-X|\to0$，那么X_n依概率收敛到X. 如果$E|X_n-X|<\varepsilon^2$，那么$P(|X_n-X|>\varepsilon)<\varepsilon$. $E|X_n|\leqslant E|X|+E|X_n-X|$和$X\in\mathcal{L}^1$蕴涵着$\{X_n\}$是$L^1$有界序列，即对所有的$n$，$E|X_n|\leqslant K<\infty$. 给定$\varepsilon>0$，取$\gamma>0$，使得$P(A)<\gamma$，则$E|X|1_A<\varepsilon/2$(这样的$\gamma$是存在的，否则，我们将取满足$P(A(n))<1/2^n$的$A(n)$，因此由博雷尔-坎泰利引理(8.3.4)得，$|X|1_{A(n)}\to0$a.s.且$E|X|1_A\geqslant\varepsilon/2$，这与控制收敛定理矛盾). 那么取足够大的$n_0$，使得对所有的$n\geqslant n_0$，$E|X_n-X|1_A\leqslant\varepsilon/2$. 取$\delta>0$，$\delta<\gamma$，使得$P(B)<\delta$，则当$n<n_0$时，$E|X_n|1_B<\varepsilon/2$.

356

(显然，可积函数的有限集是一致可积的.)那么当$n\geqslant n_0$时，$E|X_n|1_B<\varepsilon/2+\varepsilon/2=\varepsilon$. □

习题

1. 设$\{X_n,\mathcal{B}_n\}$是鞅，把它视为一个赌徒每赌完一局后的资金序列. 假设另一个赌徒B，下堵注为f_n，其中f_n是有界的$\mathcal{B}_n$可测函数，关于第$(n+1)$次结果，B的资金序列为$Y_n:=Y_1+\sum_{1\leqslant j\leqslant n}(X_{j+1}-X_j)f_j$，$n\geqslant2$，其中$Y_1\in\mathcal{L}^1$且$Y_1$是$\mathcal{B}_1$可测的. 证明：$\{Y_n,\mathcal{B}_n\}$仍是鞅.(因而，$B$没有基于过去结果的赌博系统，能在未来的赌局中获

利．以前，鞅是这样一个系统而不是序列$\{X_n\}$或$\{Y_n\}$．）

2. 设X_1，X_2，…是独立同分布随机变量且服从分布$N(0,1)$，$S_n:=X_1+\cdots+X_n$，$Y_n:=\exp(S_n-n/2)$．证明：$\{Y_n, \mathcal{B}_n\}$是一个鞅，其中$\mathcal{B}_n$是X_1，…，X_n在其上可测的最小σ-代数．［提示：参考命题9.4.2.］

3. 设Y_1，Y_2，…是独立同分布随机变量且服从分布$N(0,4)$，令$S_n:=Y_1+\cdots+Y_n$．试求常数c_n，使得$\exp(S_n-c_n)(c_1=0)$是鞅，［提示：参考上面的习题2.］

4. 在定理10.3.5中，证明：如果概率空间是具有勒贝格测度的单位区间而不是有原子（如3.5节）的概率空间，ε和δ的连续性条件蕴涵了L^1有界性．

5. 设X是任意随机变量，$0<E|X|<\infty$，$EX=0$．证明：存在一个鞅$\{X_n\}_{n=0,1,2}$，使得$X_0=0$，$X_2=X$，$P(X_1\neq 0)=1$．［提示：设$\mathcal{B}_1$是由集合$\{X<0\}$生成的．］

6. 设X，Y，Z是独立实值随机变量，$Ee^X<\infty$，$E|X|<\infty$，$EY=0$，$EZ^2<\infty$．证明：$\{X, e^X+Y+Z^2\}$是一个下鞅，其中$\mathcal{B}_1$是使X可测的最小σ-代数，并求其Doob分解．［提示：参考习题10.1.5.］

7. 求在具有勒贝格测度的$[0,1]$上，令$f_n:=a_n 1_{[0,1/n]}$，对什么样的序列a_n，$\{f_n\}$是一致可积的？

8. 设$\{X_n\}$是一个下鞅．证明：EX_n是非减数列，并举例说明X_n-EX_n不一定是鞅．［提示：考虑Doob分解的唯一性．］

357

9. 举例：鞅序列$\{X_n\}_{n\geq 1}$依概率收敛但不是几乎必然收敛的．［提示：设$\{u_n\}_{n\geq 1}$是独立同分布的，且是$[0,1]$上的均匀分布（分布＝勒贝格测度）．令$X_1\equiv 0$．如果$X_n=0$，当$0\leq u_n<1/(2n)$时，令$X_{n+1}=-1$；当$1/(2n)\leq u_n<1/n$时，令$X_{n+1}=1$；其他情况时，令$X_{n+1}=0$．如果$X_n=a\neq 0$，当$1/2^n\leq u_n\leq 1$时，令$X_{n+1}=0$，其他情况时，令X_{n+1}有一个依赖于a的适当的值．

10.4 最优停止和一致可积性

给定一个概率空间$(\Omega, \mathcal{B}, P)$，集合$T\subset\mathbb{R}$及$\mathcal{B}$的子σ-代数$\mathcal{B}_t$所构成的集族$\{\mathcal{B}_t\}_{t\in T}$，使得当$t<u$时，$\mathcal{B}_t\subset\mathcal{B}_u$，取值于$T\cup\{+\infty\}$的$\Omega$上的随机变量$\alpha$称为停时（stopping time），当且仅当对每个$t\in T$，$\{\omega: \alpha(\omega)\leq t\}\in\mathcal{B}_t$．本章中$T$是可数集，如正整数集$\mathbb{N}^+$或负整数集或$\mathbb{N}^+\cup\{\infty\}$．如果对每个$n\in T$，$X_n$是随机变量，那么集族$\{X_n\}_{n\in T}$或$\{X_n, \mathcal{B}_n\}_{n\in T}$称为适应的（adapted），当且仅当对所有的$n$，$X_n$是$\mathcal{B}_n$可测的．当涉及集族$\{X_n\}$时，在没有说明的情况下，假定$X_n$是适应的．另一方面，对任意随机变量$X_n$，$n\in T$，令$\mathcal{A}_n$是所有的$X_k(k\leq n)$在其上可测的最小$\sigma$-代数．那么$\{X_n, \mathcal{A}_n\}_{n\in T}$是适应的．如果不涉及$\sigma$-代数$\mathcal{B}_n$或没有具体化，那么可取$\mathcal{B}_n=\mathcal{A}_n$．一般地，把$T$当作时间集，$\mathcal{B}_n$中的事件是我们已知的在时刻$n$发生或不发生的事件．例如，设$\{X_n, \mathcal{B}_n\}$是一个下鞅，其中$X_n$是某商品在$n$时刻的价格，令$\alpha(\omega)$是一个人采取行动（买或卖此商品）的时刻．为了决定在n时刻买卖的行动是否发生，他需要n时刻以及n时刻以前的信息，因此α是一个停时．

例：如果$T=\{1,2,\cdots\}$，$\{X_n\}$是适应的，$y\in\mathbb{R}$，那么$\alpha(\omega):=\inf\{n: X_n(\omega)>y\}$，或者如果对所有的$n$，$X_n(\omega)\leq y$，$\alpha(\omega):=+\infty$，那么$\alpha$是一个停时．

任意常数$\alpha(\cdot)\equiv c\in T\cup\{+\infty\}$是一个停时．

任意可测函数$f\geq 0$和固定的n，$n+f(X_1,\cdots,X_n)$是一个停时．

给定一个停时α，令$\mathcal{B}_\alpha$是所有的事件$A\in\mathcal{B}$的集族，其中对所有的t，$A\cap\{\omega: \alpha(\omega)\leq t\}\in\mathcal{B}_t$．$\mathcal{B}_\alpha$称为到时刻$\alpha$为止的事件集．显然$\mathcal{B}_\alpha$是一个$\sigma$-代数．对固定的时刻，这种思想表现为$\mathcal{B}_\alpha$也是某些事件的集合，其中这些事件到$\alpha$时刻能够知道其是否发生．

358

假设$T=\{1,2,\cdots\}$，令α是一个在T中取值的停时．令$X_\alpha(\omega):=X_{\alpha(\omega)}(\omega)$．那么$\alpha$有可数个值，每一个都包含在可测集中，因此$X_\alpha$是随机变量（引理4.2.4）．事实上，很容易验证$X_\alpha$是$\mathcal{B}_\alpha$可

测的．如果 α 是有界的（如 $\alpha \leqslant m$），对所有的 n，$X_n \in \mathcal{L}^1$，那么因为 $|X_\alpha| \leqslant \max(|X_1|, \cdots, |X_m|)$，有 $X_\alpha \in \mathcal{L}^1$.

可以证明在最优停止条件下，下鞅、上鞅和鞅的性质保持不变，也就是说，如果停时是有界的，可以计算停时的值．

例： 设 $X_1, X_2, \cdots$ 是独立随机变量且对所有的 n，$P(X_n = 2^n) = P(X_n = -2^n) = 1/2$. 设 $S_n := X_1 + \cdots + X_n$. 那么如命题 10.3.2 前面的一个例子，$\{S_n, \mathcal{B}_n\}$ 是鞅．设 α 是使得 $X_n > 0$ 的最小的 n. 那么 α 是几乎必然有限的．注意 $S_\alpha = 2$a. s.．因此，尽管 1 和 $\alpha(1 \leqslant \alpha)$ 是停时，但 $E(S_\alpha \mid \mathcal{B}_1) = S_1$ 却几乎必然不成立．因而对某个无界停时，下面的定理是不成立的，幸运地是它对有界停时是成立的．

10.4.1 最优停止定理　设 $\{X_n, \mathcal{B}_n\}$ 是任意鞅序列且 α, β 是 $\mathcal{B}_n$ 的两个停时，对某个 $N < \infty$，有 $\alpha \leqslant \beta \leqslant N$. 那么 $\{X_\alpha, X_\beta; \mathcal{B}_\alpha, \mathcal{B}_\beta\}$ 是鞅，所以 $E(X_\beta \mid \mathcal{B}_\alpha) = X_\alpha$a. s.．这对下鞅或上鞅同样成立，只需分别用"$\leqslant$"或"$\geqslant$"代替"$=$"即可．换句话说，如果 $Y_1 := X_\alpha$，$Y_2 := X_\beta$，$\mathcal{S}_1 := \mathcal{B}_\alpha$，$\mathcal{S}_2 := \mathcal{B}_\beta$，$D := \{1, 2\}$，那么如果 $\{X_n, \mathcal{B}_n\}$ 有相同的性质，$\{Y_j, \mathcal{S}_j\}_{j \in D}$ 是鞅、下鞅或上鞅．

证明　如果 $A \in \mathcal{B}_\alpha$，那么对所有的 n，

$$A \cap \{\beta \leqslant n\} = (A \cap \{\alpha \leqslant n\}) \cap \{\beta \leqslant n\} \in \mathcal{B}_n,$$

因此 $A \in \mathcal{B}_\beta$. 因而 $\mathcal{B}_\alpha \subset \mathcal{B}_\beta$.

设 $\{X_n, \mathcal{B}_n\}$ 是一个下鞅．给定 $j = 1, \cdots, N$，及 $A \in \mathcal{B}_\alpha$，令 $A_j := A(j) := A \cap \{\alpha = j\}$，使得 $A_j \in \mathcal{B}_j$. 对 $j \leqslant k \leqslant N$，令

$$A_{jk} := A(j,k) := A_j \cap \{\beta = k\},$$

$$U(j,k) := \bigcup_{i \geqslant k} A_{ji} = A_j \cap \{\beta \geqslant k\},$$

$$V(j,k) := U(j,k+1) = A_j \cap \{\beta > k\}.$$

359 因为 $\{\beta > k\} = \Omega \setminus \{\beta \leqslant k\} \in \mathcal{B}_k$，所以 $V(j, k) \in \mathcal{B}_k$.

由下鞅的定义可得，$E(X_k 1_{V(j,k)}) \leqslant E(X_{k+1} 1_{V(j,k)})$. 因而

$$E(X_k 1_{U(j,k)}) \leqslant E(X_k 1_{A(j,k)}) + E(X_{k+1} 1_{V(j,k)}),$$

所以

$$E(X_k 1_{U(j,k)}) - E(X_{k+1} 1_{U(j,k+1)}) \leqslant E(X_\beta 1_{A(j,k)}).$$

从 $k = j$ 到 N 求和，左边是一个压缩和且 $U(j, N+1) = \varnothing$，因此，

$$E(X_j 1_{A(j)}) = E(X_j 1_{U(j,j)}) \leqslant E(X_\beta 1_{A(j)}).$$

从 $j = 1$ 到 N 求和可得，$E(X_\alpha 1_A) \leqslant E(X_\beta 1_A)$.

上鞅的情况是对称的（用 $-X_n$ 代替 X_n）. 如果是鞅就用等号代替不等号，证毕．□

定理 10.4.1 中鞅的情况，假设 $\alpha \leqslant \beta \equiv N$. 那么 $E(X_N \mid \mathcal{B}_\alpha) = X_\alpha$，因此 $EX_\alpha = EX_N = EX_1$. 因而在一个公平的赌局中，对一个有界停时，玩家不能由最优停止（基于停时以前的时刻获得的信息）来增加或减少他的利益期望值．例如，当你赚钱时你就停下来，令 α 是使得 $X_\alpha - X_1$ 为正值或等于一个给定的正数的时刻，如果这是可能的，你可以一直玩到给定的有限时刻 N. 那么你赔钱的概率很小，但是你赚钱的期望和赔钱的期望是均衡的，所以如果你输的话肯定会输得很多．

给定随机变量 $X_1, \cdots, X_n$，$M > 0$，事件定义为 $A(M, n) := \{\max_{1 \leqslant j \leqslant n} X_j \geqslant M\}$.

10.4.2 定理（Doob 极大不等式）　对任意下鞅序列 $\{X_n, \mathcal{B}_n\}$ 和任意的 n，$M > 0$，有

$$MP(A(M,n)) \leqslant E(X_n 1_{A(M,n)}) \leqslant E\max(X_n,0).$$

注：Doob 不等式是对马尔可夫不等式的改进，$MP(X \geqslant M) \leqslant EX^+$，其中 $X^+ := \max(X, 0)$. 如果 $X = M$ 它就变为一个等式，否则为 0，在这个意义上，马尔可夫不等式是严紧的．另一方面，Doob 不等式对事件 $\{X_j \geqslant M\}$ $(j=1, \cdots, n)$ 的并的概率给出了相同的上界，而马尔可夫不等式只对某一个事件 $\{X_n \geqslant M\}$ 给出了上界．

证明 设 α 是使得 $X_j \geqslant M(j \leqslant n)$ 的最小的 j，若不存在这样的 $j \leqslant n$，令 $\alpha := n$. 那么 α 是一个停时且 $\alpha \leqslant n$. 令 $A := A(M, n)$，那么因为对任意 $m \leqslant n$，

360

$$A \cap \{\alpha \leqslant m\} = \{\max_{j \leqslant m} X_j \geqslant M\} \in \mathcal{B}_m,$$

所以 $A \in \mathcal{B}_\alpha$. 由 $\beta \equiv n$ 时的最优停止定理（10.4.1）可知，$\{X_\alpha, X_n; \mathcal{B}_\alpha, \mathcal{B}_n\}$ 是下鞅．在 A 上，$X_\alpha \geqslant M$，因此，

$$MP(A) \leqslant E(X_\alpha 1_A) \leqslant E(X_n 1_A) \leqslant E\max(X_n,0). \qquad \square$$

一个鞅序列 $\{X_n, \mathcal{B}_n\}$ 称为右闭的（right-closable），当且仅当它能扩张到线性序集 $\{1, 2, \cdots, \infty\}$ 上，使得存在一个随机变量 $X_\infty \in \mathcal{L}^1$，且对所有的 n，$E(X_\infty \mid \mathcal{B}_n) = X_n$. 对所有的 n，设 $\mathcal{B}_\infty$ 是包含 $\mathcal{B}_n$ 的最小 σ-代数．那么设 $Y := E(X_\infty \mid \mathcal{B}_\infty)$，$Y$ 是 $\mathcal{B}_\infty$ 可测的，由定理 10.1.3 知，对所有的 n，$E(Y \mid \mathcal{B}_n) = X_n$. 因此，假设 X_∞ 是 $\mathcal{B}_\infty$ 可测的．在这种情况下，$\{X_n, \mathcal{B}_n\}_{1 \leqslant n \leqslant \infty}$ 称为右闭鞅（right-closed martingale）. 另一方面，对任意 $Y \in \mathcal{L}^1$ 和递增 σ-代数序列 $\mathcal{B}_n$，由定理 10.1.3 知，对给定 $X_n := E(Y \mid \mathcal{B}_n)$ 总给出一个鞅序列，这里给出了判断这种形式的鞅的准则．

10.4.3 定理 一个鞅序列 $\{X_n, \mathcal{B}_n\}$ 是右闭的，当且仅当 $\{X_n\}$ 是一致可积的．

注：一旦证明了此等价性，"右闭的"将不再使用，右闭鞅 $\{X_n, \mathcal{B}_n\}_{1 \leqslant n < \infty}$ 称为一致可积的．

证明 如果对所有的 n，$E|Y| < \infty$，那么对任意集 $A \in \mathcal{B}_n$，因为 $|\cdot|$ 是一个凸函数，由条件詹森不等式（10.2.7）可得，$E(|X_n| 1_A) \leqslant E(|Y| 1_A)$. $\{Y\}$ 是一致可积的，由定义及控制收敛定理可得，当 $M \to \infty$ 时，$E(|Y| 1_{|Y| > M}) \to 0$. 从一致可积性的特征函数的选择（定理 10.3.5）可得，当 $P(A) \to 0$ 时，$E(|Y| 1_A) \to 0$. 将其应用到集合 $A := \{|X_n| \geqslant M\}$ 上可得 $\{X_n\}$ 是一致可积的，因为当 $M \to \infty$ 时，

$$\sup_n P(|X_n| \geqslant M) \leqslant \sup_n E|X_n|/M \leqslant E|Y|/M \to 0,$$

反之，假设 $\{X_n\}$ 是一致可积的．设 $\mathcal{A} := \bigcup_{n \geqslant 1} \mathcal{B}_n$ 是代数（任意递增代数的并是一个代数），对任意 $A \in \mathcal{B}_n$，令 $\gamma(A) := E(X_n 1_A)$. 那么对所有的 $m \geqslant n$，由鞅的性质可得，

361

$$\gamma(A) = E(X_m 1_A) = \lim_{k \to \infty} E(X_k 1_A).$$

因而 γ 在 $\mathcal{A}$ 上有明确定义（$\gamma(A)$ 对足够大的 n 使得 $\mathcal{A} \in \mathcal{B}_n$ 且不依赖于 n）. 由一致可积性知，对任意 $\varepsilon > 0$，存在 $\delta > 0$，使得当 $A \in \mathcal{A}$，$P(A) < \delta$ 时，$|\gamma(A)| < \varepsilon$. 下面的引理是有帮助的.

10.4.4 引理 设 $\mathcal{A}$ 是一个代数，$\mathcal{S}$ 是由 $\mathcal{A}$ 生成的 σ-代数．μ 是 $\mathcal{S}$ 上有限非负测度，γ 是 $\mathcal{A}$ 上的有限可加的有界实值函数的集合．假定当 $\delta \downarrow 0$ 时，$\sup\{|\gamma(A)| : A \in \mathcal{A}, \mu(A) < \delta\} \to 0$. 那么 γ 扩张到 $\mathcal{S}$ 上的可数可加符号测度后，关于 μ 是绝对连续的，使得对某 $f \in L^1(\mu)$，

$$\gamma(B) = \int_B f \mathrm{d}\mu, \quad \forall B \in \mathcal{S}.$$

证明 如果 $A_n \downarrow \varnothing$，$A_n \in \mathcal{A}$，那么由定理 3.1.1 可得，$\mu(A_n) \downarrow 0$，因此 $\gamma(A_n) \to 0$. 因而 γ 在 $\mathcal{A}$ 上是可数可加的（由定理 3.1.1）. γ 扩张到 $\mathcal{S}$ 上的可数可加符号测度上是有界的（定理 5.6.3），仍称

为γ. 那么存在若尔当分解$\gamma=\gamma^+-\gamma^-$，其中γ^+，γ^-是有限非负测度，且对每个$C\in\mathcal{S}$，由定理 5.6.1 可得，$\gamma^+(C)=\sup\{\gamma(B): B\in\mathcal{S}, B\subset C\}$.

给定任意的$\varepsilon>0$，取$\delta>0$，使得如果$\mu(A)<\delta$，$A\in\mathcal{A}$，那么$|\gamma(A)|\leqslant\varepsilon$. 设$\mathcal{M}$为使得$\mu(B)\geqslant\delta$或$|\gamma(B)|\leqslant\varepsilon$成立的$\mathcal{S}$中的所有集合$B$的集族. 那么$\mathcal{A}\subset\mathcal{M}$. 为了证明$\mathcal{M}$是单调类，令$B_n\uparrow B$，$B_n\in\mathcal{M}$. 如果对所有的$n$，$|\gamma(B_n)|\leqslant\varepsilon$，那么由可数可加性(特别地，$\gamma^+$，$\gamma^-$在单调收敛下的连续性)得，$|\gamma(B)|\leqslant\varepsilon$. 否则对某个$n$，$\mu(B_n)\geqslant\delta$，那么$\mu(B)\geqslant\delta$，因此$B\in\mathcal{M}$. 下面假设$B_n\downarrow B$，$B_n\in\mathcal{M}$. 如果对所有足够大的$n$，有$|\gamma(B_n)|\leqslant\varepsilon$，那么$|\gamma(B)|\leqslant\varepsilon$. 否则，对某个子序列$n(k)$，$\mu(B_{n(k)})\geqslant\delta$，因此$\mu(B)\geqslant\delta$，所以有$B\in\mathcal{M}$，且$\mathcal{M}$是一个单调类. 因此$\mathcal{M}=\mathcal{S}$(定理 4.4.2).

因为对所有的$\varepsilon>0$和适当的$\delta=\delta(\varepsilon)>0$，$\mathcal{M}=\mathcal{S}$成立，那么如果$\mu(B)=0$，有$\gamma(B)=0$，且$\gamma^+(B)=0$，因此$\gamma^-(B)=0$. 因而$\gamma$，$\gamma^+$，$\gamma^-$关于测度$\mu$绝对连续. 再由 Radon-Nikodym 定理(5.5.4 和推论 5.6.2)可得出此引理. □

现在继续定理 10.4.3 的证明，运用引理 10.4.4，令X_∞是 Radon-Nikodym 导数$d\gamma/dP$. 那么对任意n和集合$A\in\mathcal{B}_n$，$\int_A X_n\,dP=\gamma(A)=\int_A X_\infty\,dP$，因此对所有的$n$，$E(X_\infty\mid\mathcal{B}_n)=X_n$. □

362

习题

1. 一个赌徒掷一枚均匀的硬币，若出现正面他就赢 1 元钱，若出现反面他就输 1 元钱. 当他或她的玩伴赢得 1 元时或者掷n次后，游戏就终止，这两种情况迟早会出现. 如果(a)$n=3$，(b)$n=5$时此玩家赢的概率是多少？如果预料输而不是赢呢？
2. 设α是一个停时，$\tau(n)$是一个有界停时非减序列，两者都是对鞅$\{X_n, \mathcal{B}_n\}_{n\geqslant1}$而言. 证明：$\alpha$对鞅$\{X_{\tau(n)}, \mathcal{B}_{\tau(n)}\}_{n\geqslant1}$也是一个停时，在此意义下，对每个$n$有$\{\alpha\leqslant\tau(n)\}\in\mathcal{B}_{\tau(n)}$.
3. 设$\{X_n, \mathcal{B}_n\}$是L^1有界鞅. 设$\mathcal{I}$是所有使得$\tau<\infty$ a.s. 的τ的集合. 证明：对$\tau\in\mathcal{I}$，所有X_τ的集合也是L^1有界的. [提示：考虑停时$\min(\tau, n)$.]
4. 设$\{X_n\}_{1\leqslant n\leqslant\infty}$是一个右闭鞅. 假定对某个$p>1$，$E|X_\infty|^p<\infty$. 令$Y:=\sup_n|X_n|$，$0<r<p$. 证明：$EY^r<\infty$. [提示：对鞅$|X_n|^p$应用极大不等式，然后对$Y^r$运用引理 8.3.6.]
5. 在习题 4 中，如果$p=1$，证明：$EY=\infty$有可能发生. [提示：令概率空间是有均匀分布(2 倍的勒贝格测度)的$(0, 1/2)$. 令$X_\infty(t)=t^{-1}|\log t|^{-3/2}$，$0<t<1/2$. 设$\mathcal{B}_n$是由区间$[j/2^n, (j+1)/2^n)(j=0, 1, \cdots)$生成的$\sigma$-代数. 运用在区间$(1/2^{n+1}, 1/2^n)$上$Y\geqslant X_n$的事实，注意：这是另一个一致可积族不被任意可积函数所控制的例子.]
6. (随机游动). 设X_1，X_2，…，是独立同分布随机变量且$X_j=\pm1$或$X_j=-1$的概率为 1/2. $S_n:=X_1+\cdots+X_n$，给定一个鞅(见命题 10.3.2 前面的例子). 令$\tau:=\min\{n: S_n=1\}$. 证明：$\tau<\infty$ a.s.. [提示：令$p=P(\tau<\infty)$. 对X_1考虑两个概率，证明$p=(p^2+1)/2$，因此$p=1$. (注意$ES_\tau=1\neq ES_1=0$，因此，这是证明定理 10.4.1 对无界停时($\alpha=1\leqslant\tau$)不成立的另一个例子.]
7. 证明：如果μ是一个有限测度，并且如同引理 10.4.4，当$\delta\downarrow0$时，$\sup\{|\gamma(A)|: A\in\mathcal{A}, \mu(A)<\delta\}\to0$，$\gamma$在$\mu$的每一个原子上是有限的，那么$\gamma$是有界的. [提示：见 3.5 节.]
8. (最优抽样). 设$(X_n, \mathcal{B}_n)$是一个下鞅，且设$\tau(1)\leqslant\tau(2)\leqslant\cdots$是非减有界停时序列(因此对每个$n$，存在某常数$M_n<\infty$使得$\tau(n)\leqslant M_n$)，证明：$(X_{\tau(n)}, \mathcal{B}_{\tau(n)})$是一个下鞅.

363

10.5 鞅和下鞅的收敛性

为证明L^1有界的鞅是几乎必然收敛的，首先给出右闭鞅收敛的证明. 收敛性不仅仅对指标集$\{1, 2, \cdots\}$，并且鞅$\{X_1, X_2, \cdots\}$伴随X_∞是闭的.

例如，对任意随机变量 X_1，X_2，…(不必是鞅或下鞅)，设$\mathcal{B}_n$是 X_1，…，X_n 在其上可测的最小 σ-代数，且$\mathcal{B}_\infty$是包含所有$\mathcal{B}_n$的最小 σ-代数．设 A 是$\mathcal{B}_\infty$中的任意集合．那么条件期望 $E(1_A \mid \mathcal{B}_n)$也可称为条件概率 $P(A \mid \mathcal{B}_n) = P(A \mid X_1, \cdots, X_n)$．下面将证明当 $n \to \infty$ 时此序列收敛到 1_A a. s.．

10.5.1 定理(Doob) *对任意右闭鞅序列* $\{X_n, \mathcal{B}_n\}_{1\leqslant n\leqslant\infty}$，*当* $n\to\infty$ *时*，X_n *几乎必然收敛到* X_∞．

证明(C. W. Lamb) 假设鞅是定义在概率空间$(\Omega, \mathcal{B}, P)$上且$\mathcal{B}=\mathcal{B}_\infty$．设 F 是使得对某个 $n=n(Y)<\infty$，Y 是$\mathcal{B}_n$可测的所有 $Y\in\mathcal{L}^1(\Omega, \mathcal{B}, P)$的集合．为了证明 F 在$\mathcal{L}^1$是稠密的，对 F 的元素逼近简单函数和示性函数 $1_A(A\in\mathcal{B})$．事件 A 的集族能近似地看作包含代数$\mathcal{A}=\bigcup\limits_{n<\infty}\mathcal{B}_n$的单调类，因此由 4.4.2 知，它和$\mathcal{B}$是相等的．

给定 $\varepsilon>0$，选择 $Y_\infty\in F$，使得 $E\mid X_\infty - Y_\infty\mid < \varepsilon^2$．设 $Y_n := E(Y_\infty \mid \mathcal{B}_n)$，那么对所有 $n\geqslant n(Y_\infty)$，有 $Y_n = Y_\infty$，因此 $Y_n\to Y_\infty$ a. s.．注意$\{X_n - Y_n, \mathcal{B}_n\}_{1\leqslant n\leqslant\infty}$是鞅．因而$\{\mid X_n - Y_n\mid, \mathcal{B}_n\}$是一个下鞅(定理 10.3.3(b)对 $f(x)\equiv\mid x\mid$)．那么由极大不等式(10.4.2)和单调收敛性(定理 4.3.2)得，

$$P(\sup_{n<\infty}\mid X_n - Y_n\mid > \varepsilon) \leqslant \sup_{n<\infty}E\mid X_n - Y_n\mid/\varepsilon \leqslant E\mid X_\infty - Y_\infty\mid/\varepsilon < \varepsilon.$$

因而

$$P\{\limsup_{n\to\infty}X_n - Y_\infty > \varepsilon\} < \varepsilon,$$
$$P\{\liminf_{n\to\infty}X_n - Y_\infty < -\varepsilon\} < \varepsilon,$$
$$P(\mid X_\infty - Y_\infty\mid > \varepsilon) < \varepsilon,$$

因此，$P(\limsup X_n - X_\infty > 2\varepsilon) < 2\varepsilon$ 且 $P(\liminf X_n - X_\infty < -2\varepsilon) < 2\varepsilon$．令 $\varepsilon\downarrow 0$，得 $X_n\to X_\infty$ a. s.． □

10.5.2 推论 *对任意的一致可积鞅*$\{X_n, \mathcal{B}_n\}_{1\leqslant n\leqslant\infty}$，*当* $n\to\infty$ *时* X_n *几乎必然收敛到某个* X_∞，*其中* X_n *满足对所有的* n，*有* $E(X_\infty\mid\mathcal{B}_n) = X_n$ *且* $E\mid X_n - X_\infty\mid\to 0$.

证明 应用定理 10.4.3、定理 10.5.1 和定理 10.3.6. □

容易看出两个停时的最大值或最小值还是一个停时．令 $a\wedge b := \min(a, b)$且 $a\vee b := \max(a, b)$．因此对任意停时 α 和任意固定的 n，$\alpha\wedge n$ 是一个停时．

364

10.5.3 引理 *对任意鞅*$\{X_n, \mathcal{B}_n\}$*和停时* α，$\{X_{\alpha\wedge n}, \mathcal{B}_{\alpha\wedge n}\}$*是一个鞅*．

证明 应用定理 10.4.1(最优停止)和命题 10.3.2. □

引理 10.5.3 说：假定一个赌徒玩一个公平的赌局，如果他连续无限地赌下去，他在 n 时刻的资金序列 X_n 是一个鞅．当必须作决定时，赌徒根据基于已有信息的某种规则，决定在 α 时刻停止．那么赌徒在 n 时刻的资金序列为 $X_{\alpha\wedge n}$并且仍是一个鞅．

10.5.4 鞅收敛定理(Doob) *设*$\{X_n, \mathcal{B}_n\}_{1\leqslant n<\infty}$*是任意* L^1 *有界鞅序列．那么* X_n *几乎必然收敛到某个随机变量* X.

在证明定理之前，我们先看一些例子．设 μ 和 ν 是同一个 σ-代数$\mathcal{S}$上的概率测度．$\mathcal{A}_n$是一个递增的 σ-代数序列，其并生成$\mathcal{S}$. 例如，设$\mathcal{S}$是[0, 1)上的 σ-代数．对 $n=1, 2, \cdots$，令$\mathcal{A}_n$是由区间 $[(k-1)/2^n, k/2^n)(k=1, 2, \cdots, 2^n)$生成的代数，假定 ν 在$\mathcal{A}_n$上关于 μ 是绝对连续的，有 Radon-Nikodym 导数 f_n，那么$\{f_n, \mathcal{A}_n\}$关于 μ 是一个鞅，它总是 L^1 有界的．如果 ν 在$\mathcal{S}$上关于 μ 是绝对连续的，有 Radon-Nikodym 导数 f，那么由定理 10.5.1 可得，f_n 几乎必然收敛到 f，且在 L^1 中．另一方面，如果 ν 关于 μ 是奇异的，定理 10.5.4 说明 f_n 仍几乎必然收敛到某处（实际上收敛到 0，习题 4 将给出证明)．

证明 给定 $M>0$，设 $\alpha(\omega)$ 是使 $|X_n(\omega)| \geq M$ 成立的最小的 n，如果这样的 n 不存在，则 $\alpha(\omega)=+\infty$. 那么 α 是一个停时. 此证明将用到下面的事实.

10.5.5 引理 $\{X_{\alpha\wedge n}, \mathcal{B}_{\alpha\wedge n}\}_{1\leq n<\infty}$ 是一致可积的鞅序列.

证明 首先，由定理 10.5.3 得，$\{X_{\alpha\wedge n}\}$ 是一个鞅. 下面证明它是一致可积的，如果 $\alpha<\infty$，令 $Z:=|X_\alpha|$；否则 $Z:=M$. 那么对所有的 n，$|X_{\alpha\wedge n}| \leq Z$ a. s.. 因此只需证明 $EZ<\infty$.（例如，如果

365

增量 $X_{n+1}-X_n$ 对所有的 n 是一致有界的，即 $|X_{n+1}-X_n| \leq K<\infty$ a. s.，那么 $Z\leq M+K$.）现在有 $E(Z1_{\{\alpha=\infty\}}) \leq M$ 且对 $\alpha\wedge n$ 和 $\beta:=n$ 运用最优停止(10.4.1)和定理 10.3.3 可得，

$$\begin{aligned}E(Z1_{\{\alpha<\infty\}}) &= E(|X_\alpha|1_{\{\alpha<\infty\}}) = E(\lim_{n\to\infty}|X_{\alpha\wedge n}|1_{\{\alpha<\infty\}})\\ &\leq \liminf_{n\to\infty} E(|X_{\alpha\wedge n}|1_{\{\alpha<\infty\}}) \quad (\text{由法图引理})\\ &\leq \sup_n E|X_{\alpha\wedge n}| \leq \sup_n E|X_n| < \infty.\end{aligned}$$

因此 $E|Z|<\infty$. □

为完成鞅的收敛性的证明，对所有的 n，在 $|X_n|<M$ 的集合上，$X_n\equiv X_{\alpha\wedge n}$，因此由引理 10.5.5 和推论 10.5.2 得，X_n 几乎必然收敛. 由极大不等式(定理 10.4.2)和单调收敛性(定理 4.3.2)知，当 $M\to\infty$ 时，$P(\text{对某个 } n, |X_n|\geq M) \leq \sup_n E|X_n|/M \to 0$. 因而 X_n 几乎必然收敛. □

例(Polya 罐子模型)：一个罐子里有一黑球和一红球. 重复地从罐子里随机地取球，然后用另一个相同颜色的球来替换，n 次替换后的分布是什么？设 B_n 是 n 次替换后黑球的数目，R_n 是 n 次替换后红球的数目. 那么 $B_n/(B_n+R_n)$ 是一个鞅(下面的习题 2)，显然，它是一致可积的，因此，它几乎必然收敛且在 L^1 中，它的极限不是一个常数.

为把收敛定理推广到下鞅或上鞅上去，要利用下面关于 Doob 分解(10.3.4)的事实.

10.5.6 引理 如果 $(X_n, \mathcal{B}_n)$ 是一个 L^1 有界的下鞅，那么在 Doob 分解 $X_n=Y_n+Z_n$ 中，递增过程 Z_n 是一致可积的，因此 Z_n 和 Y_n 是 L^1 有界的. 如果 X_n 是一致可积的，那么 Z_n 和 Y_n 也是一致可积的. 对于上鞅也有类似的结论.

证明 我们有 $Z_1=0\leq Z_2\leq Z_3\leq\cdots$ 且 $EZ_n=EX_n-EY_n=EX_n-EY_1=EX_n-EX_1$. 因而 $\sup_n EZ_n<\infty$ 且由单调收敛性可得，$E\sup_n Z_n<\infty$，则 Z_n 可由一个可积函数控制，因此是一致可积的. 引理的其余部分显然成立. □

10.5.7 下鞅和上鞅收敛定理 任意 L^1 有界的下鞅或上鞅几乎必然收敛；如果是一致可积的，它在 L^1

366

中也收敛.

证明 对 Doob 分解，由引理 10.5.6 的证明知，$Z_n(\omega)$ 关于 n 是几乎必然非减且几乎必然收敛到一个有限可积的变量. 那么由定理 10.5.4 知，$(Y_n, \mathcal{B}_n)$ 是 L^1 有界的鞅且几乎必然收敛. 由推论 10.5.2 知，对 Y_n 和单调或控制收敛的 Z_n 可得一致可积的情况. □

例如，对任意 L^1 有界的鞅 $\{X_n\}$，其几乎必然收敛，仅因为是 $\{X_n\}$ 收敛或者作为下鞅收敛的一种特殊情况，下鞅 $\{|X_n|\}$ 也是几乎必然收敛.

最后，对非负上鞅，最优停止(定理 10.4.1)能被改进为停时仅需有限而不需要有界.

***10.5.8 定理** 设 $(X_n, \mathcal{B}_n)$ 是一个上鞅且对所有的 n 和 ω，$X_n(\omega)\geq 0$. 设 σ 和 τ 是两个停时且 $\sigma\leq\tau$. 那么

$$E(X_\tau 1_{\{\tau<\infty\}} \mid \mathcal{B}_\sigma) \leq X_\sigma 1_{\{\sigma<\infty\}}.$$

证明 对任意 n，由最优停止(定理 10.4.1)知，$EX_{\tau\wedge n}\leq EX_1$. 然后由法图引理得，$EX_\tau 1_{\{\tau<\infty\}}\leq$

$EX_1<\infty$. 同理，$EX_\sigma 1_{\{\sigma<\infty\}}<\infty$. 在$\mathcal{B}_\sigma$中的集合上，$\sigma=\infty$，$\tau=\infty$，因此，上述不等式的两边都为0，故它是成立的.

对任意的 $m=1, 2, \cdots$, 和 $A\in\mathcal{B}_\sigma$，下面证明

$$A^{(m)}:=A\cap\{\sigma\leqslant m\}\in\mathcal{B}_{\sigma\wedge m}.$$

首先，如果 $m\leqslant j$，那么 $A^{(m)}\cap\{\sigma\wedge m\leqslant j\}=A^{(m)}\in\mathcal{B}_m\subset\mathcal{B}_j$，而当 $m>j$ 时，$A^{(m)}\cap\{\sigma\wedge m\leqslant j\}=A\cap\{\sigma\leqslant j\}\in\mathcal{B}_j$. 因而对任意 $m\leqslant n$，对 $\alpha:=\sigma\wedge m\leqslant\beta:=\tau\wedge n$ 应用最优停止(定理 10.4.1)得，

$$\begin{aligned}E1_AX_\sigma 1_{\sigma\leqslant m}&=E1_AX_{\sigma\wedge m}1_{\sigma\leqslant m}\geqslant E1_AX_{\tau\wedge n}1_{\sigma\leqslant m}\\&\geqslant E1_AX_{\tau\wedge n}1_{\tau\leqslant n}1_{\sigma\leqslant m}\\&=E1_AX_\tau 1_{\tau\leqslant n}1_{\sigma\leqslant m}.\end{aligned}$$

令 $n\to\infty$，$1_{\tau\leqslant n}$递增到 $1_{\tau<\infty}$. 所有其他因子是非负单调收敛的，即可得

$$E1_AX_\tau 1_{\sigma\leqslant m}1_{\tau<\infty}\leqslant E1_AX_\sigma 1_{\sigma<\infty}.$$

同理，$1_{\sigma\leqslant m}$递增到 $1_{\sigma<\infty}$，因此再次用单调收敛定理，因为 $\sigma\leqslant\tau$，$1_{\sigma<\infty}1_{\tau<\infty}=1_{\tau<\infty}$，因此得到

$$E1_AX_\tau 1_{\tau<\infty}\leqslant E1_AX_\sigma 1_{\sigma<\infty},$$

故这个结论成立. □ 367

例：设 $X_1, X_2, \cdots$是独立实值随机变量均值为1且 $T_n:=X_1X_2\cdots X_n$. 令$\mathcal{B}_n$是 $X_1, X_2, \cdots$在其上可测的最小 σ-代数，那么因为 X_{n+1}和$\mathcal{B}_n$是独立的，$E(X_{n+1}\mid\mathcal{B}_n)=EX_{n+1}=1$. 因而定理 10.1.9 蕴涵着 $E(T_{n+1}\mid\mathcal{B}_n)=T_n$，因此，$\{T_n, \mathcal{B}_n\}$是一个鞅(命题 10.3.2). 如果 X_j 是同分布的且 $X_j>0$a. s.，令 $Y_j:=\log X_j$，那么由詹森不等式(10.2.6)和 10.2 节习题 14 知，除了 $X_1=1$a. s. 外，$EY_1<0$，如果 $-\infty<EY_1<0$，那么$(Y_1+\cdots+Y_n)/n\to EY_1$a. s.，因此 $Y_1+\cdots+Y_n\to-\infty$ a. s.，如果 $EY_1=-\infty$，取足够大的 M，使得 $E\max(Y_j, -M)<0$ 且 $Y_1+\cdots+Y_n\to-\infty$. 因而 $T_n\to 0$a. s.. 因为 $ET_n\equiv 1$，所以这样的$\{T_n\}$给出了 L^1 有界的正鞅在 L^1 中不收敛的例子.

例：这个例子中的鞅和停时是常见的. 令 D_n 是独立随机变量且对 $n=0, 1, 2, \cdots$，有 $P(D_n=2^n)=P(D_n=-2^n)=1/2$. 令 $X_n:=\sum_{j=0}^{n}D_j$，因为 $ED_j=0$，那么 X_n 可组成一个鞅. 对每个 n，取 σ-代数$\mathcal{A}_n$，且 $D_0, \cdots, D_n$ 对此最小 σ-代数是可测的或等价地 $X_0, X_1, \cdots, X_n$ 在其上是可测的.

设 τ 是使得 $X_n>0$ 或等价地 $D_n>0$(因为对所有的 n，$-1-2-\cdots-2^{n-1}+2^n=1>0$)成立的最小的 n. 那么 $P(\tau\geqslant n)=1/2^n$，因此 τ 是几乎必然有限的，并且因为 $X_\tau\equiv 1$，我们有 $EX_\tau=1$. 然而对每个 n，由最优停止定理 10.4.1 知，$EX_{\tau\wedge n}=0$. 事实上，$X_{\tau\wedge n}=1$ 的概率为 $1-2^{-n-1}$，$X_{\tau\wedge n}=-2^{n+1}+1$ 的概率为 2^{-n-1}. $X_{\tau\wedge n}$是 L^1 有界的，且由引理 10.5.3 知，$(X_{\tau\wedge n}, \mathcal{A}_{\tau\wedge n})$是鞅，由定理 10.5.4 知，$X_{\tau\wedge n}$几乎必然收敛到1. 由推论 10.5.2 知，它们不依期望收敛，故可直接验证变量 $X_{\tau\wedge n}$ 不是一致可积的. 最初的鞅$\{X_n\}$不是 L^1 有界的.

习题

1. 设 $X_1, X_2, \cdots$是独立同分布随机变量且 $X_1\geqslant 0$，$1<EX_1<\infty$，$T_n:=X_1X_2\cdots X_n$，$\mathcal{B}_n$是 $X_1, X_2, \cdots, X_n$ 在其上可测的最小 σ-代数.
   (a)证明：$\{T_n, \mathcal{B}_n\}$是一个下鞅.
   (b)求出 T_n 的 Doob 分解.
   (c)如果 $P(X_1=0)>0$，证明：尽管$\{T_n\}$不是 L^1 有界的，但 $T_n\to 0$a. s..

368

2. 在 Polya 罐子模型中(引理 10.5.6 前面的例子)证明：$B_n/(B_n+R_n)$ 是一个鞅．它几乎必然收敛到一个非常数的极限．
3. 证明或举反例：对每个鞅 $\{X_n, \mathcal{B}_n\}$ 和停时 α，$\{X_{\alpha\wedge n}, \mathcal{B}_n\}$ 是一个鞅．注意：由引理 10.5.3 知，$\{X_{\alpha\wedge n}, \mathcal{B}_{\alpha\wedge n}\}$ 是一个鞅．
4. 设 $\mathcal{A}_n$ 是一个递增的 σ-代数序列，它们的并生成 σ-代数 $\mathcal{S}$. 设 μ 和 ν 是 $\mathcal{S}$ 上的概率测度且 ν 在 $\mathcal{S}$ 上关于 μ 是奇异的，但在每个 $\mathcal{A}_n$ 上关于 μ 是绝对连续的且有 Radon-Nikodym 导数 f_n. 证明：对每个 μ，$f_n \to 0$a. s.．[提示：如果 $f_n \to f$a. s.，由法图引理可证明对每个 $A\in\mathcal{S}$，$\int_A f\mathrm{d}\mu \leqslant \nu(A)$．]
5. 设 $X_1, X_2, \cdots$ 是独立实值随机变量，且对所有的 n，$EX_n=0$，$S_n := X_1+\cdots+X_n$. 设 $\mathcal{B}_n$ 是 $X_1, \cdots X_n$ 在其上可测的最小 σ-代数．那么 $\{S_n, \mathcal{B}_n\}$ 是一个鞅(见命题 10.3.2 前的例子).
   (a) 举例说明：S_n 几乎必然收敛，但非 L^1 有界．
   (b) 如果对所有的 j，$|X_j| \leqslant 2$a. s. 且 S_n 几乎必然收敛，证明：S_n 在 L^2 中收敛．[提示：用三级数定理(9.7.3).]
6. (Wald 等式)．设 $X_1, X_2, \cdots$ 是独立实值随机变量且 $E|X_1|<\infty$，S_n，$\mathcal{B}_n$ 如习题 5 中所定义的．那么 $\{S_n, \mathcal{B}_n\}$ 是下鞅或上鞅，并且依赖于 EX_1 的符号．设 τ 关于 $\{S_n\}$ 是一个停时．如果 $E\tau<\infty$．证明：$ES_\tau = E\tau EX_1$．[提示：$ES_\tau = \sum_{n\geqslant 1} P(\tau=n)\sum_{i=1}^{n} E(X_i \mid \tau=n) = \sum_{i\geqslant 1} E(X_i \mid \tau\geqslant i)\times P(\tau\geqslant i)$．现在 $\{\tau\geqslant i\} = \{\tau > i-1\}\in\mathcal{B}_{i-1}$，且 X_i 与这个事件独立，$\sum_{i\geqslant 1} P(\tau\geqslant i) = E\tau$，可以考虑用 $|X_i|$ 代替 X_i 来进行和交换．]
7. 证明：对任意随机变量 Y，$E|Y|^p = \int_0^\infty pt^{p-1}P(|Y|\geqslant t)\mathrm{d}t$．[提示：除非你能用分部积分证明，否则，首先对简单函数 Y 证明，然后取极限．]
8. 设 X 和 Y 是两个非负的随机变量，其满足对每个 $t>0$，$P(Y\geqslant t) \leqslant t^{-1}\int_{Y\geqslant t} X\mathrm{d}P$．对任意 $p>1$，$\|f\|_p := \left(\int |f|^p \mathrm{d}P\right)^{1/p}$ 且 $q^{-1}+p^{-1}=1$，证明：$\|Y\|_p \leqslant q\|X\|_p$．[提示：运用习题 7 的结论，交换积分次序，并用 Rogers-Hölder 不等式(5.1.2).]
9. 给定一个非负下鞅 $\{X_n, \mathcal{B}_n\}$，设 $X_n^* := \max_{j\leqslant n} X_j$ 且 $X^* := \max_{j\leqslant n} X_j$. 证明：对任意 $p>1$ 且 $q^{-1}+p^{-1}=1$，$\|X^*\|_p \leqslant q\cdot\sup_n \|X_n\|_p$．[提示：用上面的习题和定理 10.4.2；先对有限的 n 给出证明．]

369

10. 设 $X_1, X_2, \cdots$ 是独立随机变量且 $EX_j=0$，有有限方差 σ_j^2. 令 $S_n = X_1+\cdots+X_n$.
    (a) 证明：$\{S_n^2\}$ 是一个下鞅．
    (b) 求 $\{S_n^2\}$ 的 Doob 分解．
    (c) 如果 $\sum_j \sigma_j^2 < \infty$，证明：鞅 S_n 几乎必然收敛(给一个三级数定理的鞅的证明，9.7 节).
11. 在习题 10 中假设对所有的 j，$\sigma_j^2=\sigma^2$. 设 T 是关于 $\{S_n\}$ 的停时且 $ET<\infty$．证明：$ES_T^2 = \sigma^2 ET$. [提示：证明结论对 $T\wedge n$ 成立，令 $n\to\infty$．如果 $T\leqslant n$，对 n 用归纳法，参考习题 6 的提示．]
12. 设 X_j 是独立的且 $P(X_j=1)=P(X_j=-1)=1/2$，S_n 与习题 10 和习题 11 中一样是简单随机游动，对正整数 m 和 n，设 T_{mn} 是使得 $S_k=-m$ 或 $S_k=n$ 成立的最小的 k. 运用习题 11 的结论计算 ET_{mn}. [提示：对 $t=T_{mn}$ 证明 $ES_t=0$ 并求 $\mathcal{L}(S_t)$.]
13. (a) 如果 $Y_n \to Y$a. s.，所有的 Y_n 和 Y 是相互独立的，证明：Y 几乎必然是一个常数．
    (b) 在定理 10.5.1 前面的例子中，如果 A 是尾事件，在这个意义上，对所有的 n，其和 $X_1, \cdots, X_n$ 是独立的．证明：$P(A)=0$ 或 1，对改进了的科尔莫戈罗夫 0－1 律(8.4.4)给出一个鞅的证明．

*10.6 逆鞅和逆下鞅

理解逆鞅的一种方法是取一个序列 $(X_n, \mathcal{B}_n)_{n\geqslant 1}$，其中 σ-代数 $\mathcal{B}_n$ 随着 n 的递增而递减．这里建

立这种序列的方法是序列的指标集由负整数按它们通常的序 $n=-1,\ -2,\ \cdots$ 来代替．此序列可记作 $\{X_n,\ \mathcal{B}_n\}_{n\leqslant -1}$．将10.3节开始定义的鞅、下鞅或上鞅中的指标集用负整数代替，则分别称为逆鞅、逆下鞅或逆上鞅．显然，逆鞅对所有的 $n=-1,\ -2,\ \cdots$，满足 $X_n=E(X_{-1}\mid \mathcal{B}_n)$．从这个意义上来讲，它像一个右闭鞅，是一个随机变量关于 σ-代数序列 $\mathcal{B}_n$ 的条件期望的集合，并由定理10.4.3的证明知，它是一致可积的．但当 $n\to-\infty$ 时，逆鞅和逆下鞅的收敛性比较有用．

例：设 $X_1,\ X_2,\ \cdots$ 是独立同分布的且 $E\mid X_1\mid<\infty$，$S_n=X_1+\cdots+X_n$．令 $Y_{-n}:=S_n/n$，对所有的 $k\geqslant n$，$\mathcal{B}_{-n}$ 是 S_k 在其上可测的最小 σ-代数．那么 $\{Y_n,\ \mathcal{B}_n\}_{n\leqslant -1}$ 是一个逆鞅，它的几乎必然收敛性和强大数定律是等价的(见下面的习题1)． 370

10.6.1 定理　*对任意的逆鞅* $(X_n,\ \mathcal{B}_n)_{n\leqslant -1}$，$X_n$ *几乎必然收敛，且在* $\mathcal{L}^1$ *中当* $n\to-\infty$ *时收敛到* $X_{-\infty}:=E(X_{-1}\mid \mathcal{B}_{-\infty})$，*其中* $\mathcal{B}_{-\infty}:=\bigcap_n \mathcal{B}_n$．

证明　首先假设 $EX_{-1}^2<\infty$．如果 X_n 定义在概率空间 $(\Omega,\ \mathcal{S},\ P)$ 上，那么 X_{-1} 属于希尔伯特空间 $H:=L^2(\Omega,\ \mathcal{S},\ P)$．由定理10.2.9知，其他变量 $X_n=E(X_{-1}\mid \mathcal{B}_n)$ 也属于 H，且算子 $T_n(X):=E(X\mid \mathcal{B}_n)$ 是 H 映上到线性子空间 $H_n:=L^2(\Omega,\ \mathcal{B}_n,\ P)$ 上的正交投影．令 $D_n:=X_n-X_{n-1}=E(X_{-1}\mid \mathcal{B}_n)-E(X_{-1}\mid \mathcal{B}_{n-1})$，$n=-1,\ -2,\ \cdots$．因为 $X_n\in H_n$，且由帕塞瓦尔等式(定理5.4.1)得，D_n 与 H 中所有的元素都正交，且

$$\|X_{-1}\|^2=\|D_{-1}\|^2+\cdots+\|D_{n+1}\|^2+\|X_n\|^2,\quad n=-2,-3,\cdots,$$

因此 $\sum_{n\leqslant -1}\|D_n\|^2$ 收敛．由Riesz-Fischer定理5.4.5(如果 $\|D_n\|>0$，$x_n:=\|D_n\|$，$e_n:=D_n/\|D_n\|$，)得，正交元素的级数 $D_{-1}+D_{-2}+\cdots$ 是收敛的．因而当 $n\to-\infty$ 时，序列 $X_n=X_{-1}-D_{-1}-\cdots-D_{1+n}$ 在 L^2 中收敛到某个 Y．令 $H_{-\infty}$ 为所有 H_n 的交．为了得到 $H_{-\infty}=L^2(\Omega,\ \mathcal{B}_{-\infty},\ P)$，需要下面的引理．

10.6.2 引理　*设* $\{\mathcal{B}_n\}_{n\leqslant -1}$ *(如定理10.6.1中所述)是一个* σ-*代数．如果* B *是一个可测集，使得对每个* n，*对某个* $A_n\in\mathcal{B}_n$，$P(A_n\Delta B)=0$．*那么对某个* $A\in\mathcal{B}_{-\infty}$，$P(A\Delta B)=0$．

证明　令 $A:=\limsup_{n\to-\infty}A_n:=\bigcap_m\bigcup_{n\leqslant m}A_n$．那么由于 $m\downarrow-\infty$ 时，$C_m:=\bigcup_{n\leqslant m}A_n$ 是递减的，且 $C_m\in\mathcal{B}_m$，$A\in\mathcal{B}_{-\infty}$．对每个 m，$P(C_m\setminus B)\leqslant\sum_{n\leqslant m}P(A_n\setminus B)=0$ 且 $P(B\setminus C_m)\leqslant P(B\setminus A_m)=0$，那么由单调收敛性可得 $P(A\Delta B)=0$．

□

10.6.3 引理　$H_{-\infty}=L^2(\Omega,\ \mathcal{B}_{-\infty},\ P)$，*也就是说，如果* $F\in L^2(\Omega,\ P)$，*等价类* F *在* $H_{-\infty}$ *中当且仅当对某个* $g\in\mathcal{L}^2(\Omega,\ \mathcal{B}_{-\infty},\ P)$，$g\in F$．

证明　显然，对每个 n，$\mathcal{B}_{-\infty}$ 可测的 $\mathcal{L}^2$ 函数在 H_n 中，因此在 $H_{-\infty}$ 中．反之，如果 $f\in H_{-\infty}$，对每个有理数 t，$\{f>t\}$ 几乎必然等于 $\mathcal{B}_n$ 的一个集合，则由引理10.6.2知，其也几乎必然等于一个集合 $C_t\in\mathcal{B}_{-\infty}$．令 $h(\omega):=\sup\{t\in\mathbb{Q}:\ \omega\in C_t\}$，其中 $-\infty\leqslant t\leqslant\infty$．那么对任意 $y\in\mathbb{R}$，$\{h>y\}=\cup\{C_t:\ t>y\}$，因此，$h$ 是 $\mathcal{B}_{-\infty}$ 可测的．在 $\bigcup_{t\in\mathbb{Q}}\{f>t\}\Delta C_t$ 外，有一个零概率集，$h=f$，引理10.6.3得证． 371

□

现在继续定理10.6.1的证明，当 $n\to-\infty$ 时，X_n 依 L^2 收敛到 $\mathcal{L}^2(\Omega,\ \mathcal{B}_{-\infty},\ P)$ 中的函数 Y．对任意集合 $T\in\mathcal{B}_{-\infty}$ 和所有的 n，$E(X_n1_T)=E(X_{-1}1_T)$，因此这些项也等于 $E(Y1_T)$．由条件期望的唯一性(定理10.1.1)得，$Y=E(X_{-1}\mid \mathcal{B}_{-\infty})$．由Cauchy-Bunyakovsky-Schwarz不等式知，对任意的 f，$E\mid f\mid=E\mid f\mid\cdot 1\leqslant(Ef^2)^{1/2}$，因此 L^2 收敛蕴涵着 L^1 收敛．令 $X_{-\infty}:=Y$．

$\{X_n - X_{-\infty}\}_{n\leqslant -1}$也是一个(逆)鞅．因而对任意关于 n 取值的有限集 F，最大数为 m，且 $\varepsilon>0$，由 Doob 极大不等式(10.4.2)得，

$$Pr\{\max_{k\in F} X_k - X_{-\infty} > \varepsilon\} \leqslant E(X_m - X_{-\infty})^+/\varepsilon.$$

那么由单调收敛性，令 F 递增到所有负整数组成的集合，可得到

$$Pr\{\sup_{k\leqslant m} X_k - X_{-\infty} > \varepsilon\} \leqslant E(X_m - X_{-\infty})^+/\varepsilon,$$

那么当 $m\to\infty$，它收敛到0. 对$\sup_{k\leqslant m} X_{-\infty} - X_k$ 有一个对称不等式．令 $\varepsilon\downarrow 0$，我们将得到当 $n\to-\infty$ 时，X_n 几乎必然收敛到 $X_{-\infty}$．

现在假设 X_{-1}不必在$\mathcal{L}^2$中．给定 $\varepsilon>0$，存在 $Y_{-1}\in\mathcal{L}^2$且有 $E|X_{-1}-Y_{-1}|<\varepsilon^2$；例如，对 $|X_{-1}|\leqslant M$，令 $Y_{-1}=X_{-1}$，否则，对足够大的 M 有 $Y_{-1}=0$. 令 $Y_n:=E(Y_{-1}|\mathcal{B}_n)$，$n=-2, 3, \cdots$ 那么$\{Y_n\}$和$\{X_n-Y_n\}$是两个逆鞅．对 Y_n 结论是成立的．对 X_n-Y_n，我们再次运用极大不等式得，

$$Pr\{\sup_n |X_n - Y_n| > \varepsilon\} \leqslant E|X_{-1}-Y_{-1}|/\varepsilon < \varepsilon.$$

因而 $\Pr(\limsup X_n - \liminf X_n > 2\varepsilon) < \varepsilon$. 令 $\varepsilon\downarrow 0$，可以得到当$n\to-\infty$ 时，X_n 几乎必然收敛．因为它是一致可积的，序列依$\mathcal{L}^1$收敛．由条件期望的唯一性知，其极限是 $E(X_{-1}|\mathcal{B}_{-\infty})$，定理 10.6.1 得证．□

对任意仅有有限多个指数的鞅$\{(Y_j, \mathcal{B}_j)\}_{1\leqslant j\leqslant n}$，对 $j=1, \cdots, n$，令 $X_{j-n-1}:=Y_j$，对 $m\leqslant n$，令 $X_m:=X_{-n}$. 那么当 $m\to-\infty$ 时，$\{X_m\}_{m\leqslant -1}$是一个逆鞅收敛的例子．因此，鞅和逆鞅的本质区别是仅对无穷指标集而言．

372

为了更好地解决逆鞅的问题，进而解决鞅的问题，需要用到 Doob 分解．

10.6.4 定理(Doob 分解和逆下鞅的收敛性) *设$\{X_n, \mathcal{B}_n\}_{n\leqslant -1}$是一个逆鞅，即$\mathcal{B}_n$是一个 σ-代数且有$\cdots\subset\mathcal{B}_{-2}\subset\mathcal{B}_{-1}$，$X_n$ 是$\mathcal{B}_n$可测的且对每个 n 是可积的，$E(X_{n+1}|\mathcal{B}_n)\geqslant X_n$，$n=-2, -3, \cdots$. 假定 $K:=\inf_n EX_n > -\infty$，那么存在一个分解 $X_n\equiv Y_n+Z_n$，其中 Z_n 是$\mathcal{B}_{n-1}$可测的，$\{Y_n, \mathcal{B}_n\}$是一个逆鞅，且几乎必然有 $Z_{-1}\geqslant Z_{-2}\geqslant\cdots$，当 $n\to-\infty$ 时，$Z_n\downarrow 0$a. s.. 因此，当 $n\to-\infty$ 时，X_n 几乎必然收敛且在$\mathcal{L}^1$中．*

证明 令 $D_n:=E(X_n|\mathcal{B}_{n-1})-X_{n-1}$，$n=-1, -2, \cdots$. 由假设知，$D_n\geqslant 0$a. s.. 那么可有 $\sum_{n\leqslant -1} ED_n = \sum_{n\leqslant -1} E(X_n - X_{n-1}) = EX_{-1}-K$；注意到，当 $n\to-\infty$ 时，$EX_n\downarrow K$. 因为和是有限的，由单调收敛定理4.3.2 知，级数 $\sum D_n$ 收敛到一个可积函数．

对每个 n，令 $Z_n := \sum_{k\leqslant n} D_k$. 那么当 n 递增时，Z_n 是几乎必然非减的，且当 $n\to-\infty$ 时，$Z_n\downarrow 0$a. s.. 令 $Y_n:=X_n-Z_n$. 为证明 Y_n 是一个(逆)鞅，我们有，

$$\begin{aligned} E(Y_n|\mathcal{B}_{n-1}) &= E(X_n - Z_n|\mathcal{B}_{n-1}) = D_n + X_{n-1} - \sum_{k\leqslant n} D_k \\ &= X_{n-1} - \sum_{k<n} D_k = X_{n-1} - Z_{n-1} = Y_{n-1}, \end{aligned}$$

故分解得到证明．由单调收敛或控制收敛得，Z_n 在$\mathcal{L}^1$中收敛，且由定理 10.6.1 知，Y_n 在$\mathcal{L}^1$中几乎必然收敛，故 X_n 在$\mathcal{L}^1$也几乎必然收敛．□

习题

1. 如同 10.2 节的习题4，设 $X_1, X_2, \cdots$是独立同分布实值随机变量且 $E|X_1|<\infty$. 令 $S_n:=X_1+\cdots+X_n$. 对

$n=1,2,\cdots$，令 $T_{-n}:=S_n/n$. $\mathcal{B}_{-n}$是 $S_k(k\geq n)$在其上可测的最小 σ-代数．证明：$\{T_k,\mathcal{B}_k\}_{k\leq -1}$是一个逆鞅且 $\{|T_k|,\mathcal{B}_k\}_{k\leq -1}$是一个逆下鞅．

2.（延拓）．设 Y 是一个随机变量且与所有的 X_j 独立，$E|Y|<\infty$．令 $U_j:=Y+T_j$，其中 $j=-1,-2,\cdots$. $\mathcal{C}_k$是 $U_j(j\leq k)$在其上可测的最小 σ-代数．证明：$\{U_k,\mathcal{C}_k\}_{k\leq -1}$是一个逆鞅，且当 $k\to-\infty$ 时，求它的极限．

373

令$\mathcal{C}_{-\infty}:=\bigcap_{k\leq -1}\mathcal{C}_k$，$Y$ 是否几乎必然等于一个$\mathcal{C}_{-\infty}$可测变量？

3. 证明：对所有的 n，任意逆鞅$\{X_n\}$($X_n\leq 0$a. s.)一致可积．
4. 令$\{\mathcal{B}_n\}_{n\leq -1}$是一个 σ-代数且$\mathcal{B}_{n-1}\subset\mathcal{B}_n(n\leq -1)$. B 是一个可测集，使得对每个 n，对某个 $A_n\in\mathcal{B}_n$，当 $n\to-\infty$ 时，$P(A_n\Delta B)\to 0$. 证明：对某个 $A\in\mathcal{B}_{-\infty}$，$P(A\Delta B)=0$.（这是对引理 10.6.2 的改进．）
5. 证明：在定理 10.6.4 中，满足所述性质的分解 $X_n=Y_n+Z_n$ 是唯一的．[提示：首先证明 Z_n-Z_{n-1}是唯一的．]
6. 证明：如果 Z_n 仅是$\mathcal{B}_n$可测的而不是$\mathcal{B}_{n-1}$可测的，那么分解 $X_n=Y_n+Z_n$ 不是唯一的．

*10.7 次加性遍历定理和超加性遍历定理

本节将证明遍历定理(8.4 节)的一种扩张，由保测变换的逐次幂来代替变量的部分和，我们仅能得到与部分和有关的单边不等式．

设$(X,\mathcal{A},P)$是一个概率空间，T 为 X 映上到其自身的保测变换．那么 X 上实值可积的随机变量序列称为有*超加性*(superadditive)，当且仅当对任意正整数 m 和 n，$f_{m+n}\geq f_m+f_n\circ T^m$a. s.．若序列的逆向不等式成立，则该序列是*次加的*(subadditive)．如果两者都成立，则称为*可加的*(additive)．显然，序列$\{f_n\}$是次加的当且仅当$\{-f_n\}$是超加的，因此，有关次加性序列的结论和超加性序列的结论是等价的．

如同遍历定理(8.4.1)中所述，设f是关于 P 可积的且 $S_n:=\sum_{0\leq j<n}f\circ T^j$，其中 T^0定义为单位元．因此$f\circ T^0\equiv f$. 那么容易得出$\{S_n\}_{n\geq 1}$是可加的．反之，对任意可加序列$\{f_n\}$，令 $f:=f_1$. 那么由归纳法可以得出，对所有的 n，$S_n=f_n$a. s.，因此，每个可加序列的形式为$\{S_n\}$.

例：给定一个$\mathbb{R}$上的法则 P，如 8.2 节那样，取$(\mathbb{R},P)$的可数积，令 X_1，X_2，$\cdots$为基坐标，因此，X_j 是关于法则 P 的独立同分布，如 8.4 节所示令 T 是移位变换 $T(\{X_n\}):=\{X_{n+1}\}$. 那么部分和$S_n:=X_1+\cdots+X_n$ 是一个可加序列，且$\{|S_n|\}_{n\geq 1}$是次加的．

设$\mathcal{A}_{\mathrm{inv}(T)}$是$\mathcal{A}$中不变集 B 的 σ-代数，$B=T^{-1}(B)$．令 E_I 表示关于$\mathcal{A}_{\mathrm{inv}(T)}$的条件期望．回顾 8.4 节，如果 T 是遍历的，当它是像最后的例子中移位变换(由定理 8.4.5)那样时，此不变集的概率为 0 或 1，所以对任意可积随机变量 g，$E_Ig=Eg$.

374

下面是关于超加性序列的一个重要的收敛定理：

10.7.1 定理(J. F. C. Kingman)　*对任意可积随机变量的超加性序列$\{f_n\}$，当 $n\to\infty$ 时，f_n/n 几乎必然收敛到 $\gamma:=\sup_n E_I(f_n)/n\leq+\infty$，其中 γ 是可积的当且仅当$\sup_n Ef_n/n<+\infty$，如果 γ 是可积的，那么 f_n/n 收敛到 γ 且 γ 也在 L^1 中．*

证明(J. Neveu)　令 $f_0:=0$，这与超加性是一致的．对任意的f，令$f^+:=\max(f,0)$，$f^-:=f^+-f$. 令

$$F_n:=\sum_{j=0}^{n-1}f_1\circ T^j,\quad H_n:=\sum_{j=0}^{n-1}f_1^+\circ T^j.$$

那么对所有的 n，$\{F_n\}_{n\geq 1}$是可加的且$f_n\geq F_n$，由超加性并对 n 进行归纳：

$$f_{n+1}\geq f_n+f_1\circ T^n\geq F_n+f_1\circ T^n=F_{n+1}.$$

令 $G_n:=f_n-F_n$. 那么 $G_n\geqslant 0$ 且 $\{G_n\}$ 是超加的. 令 $J_n:=H_n-F_n=\sum_{0\leqslant j<n}f_1^-\circ T^j$, 那么 $\{H_n\}$ 和 $\{J_n\}$ 是非负可加的且 $f_n\equiv G_n+H_n-J_n$. 对可加序列 $\{J_n\}$, 对所有的 n, 有 $E_I(J_n)/n=E_If_1^-$, 所以 $\sup_n E_I(J_n/n)=\inf_n E_I(J_n/n)$ a. s.. 因而如果 J_n/n 几乎必然收敛到 $\gamma:=\gamma(\{J_n\})$, 那么也有 $-J_n/n\to\gamma(\{-J_n\})=-\gamma(\{J_n\})$ a. s., 且对 $\{H_n\}$ 也有类似的结果. 因此, 如果对 $\{G_n\}$, $\{H_n\}$ 和 $\{J_n\}$ 定理成立, 那么因为 $\gamma(\{f_n\})=\gamma(\{G_n\})+\gamma(\{H_n\})-\gamma(\{J_n\})$, 故对 $\{f_n\}$ 结论也成立. 所以可以假定 $f_n\geqslant 0$.

证明的主要部分是基于里斯(F. Riesz)证明伯克霍夫(Birkhoff)遍历定理的方法.

10.7.2 引理(Riesz) 设 $u_1, \cdots, u_n$ 是任意实数且 $s_j:=u_1+\cdots+u_j$. 令 $v_j:=v(j, n):=\max_{k:j\leqslant k\leqslant n}s_k-s_j$. 那么

$$\sum_{j=0}^{n-1}u_{j+1}1_{\{v(j,n)>0\}}\geqslant 0.$$

注: 如果 $u_j+\cdots+u_{j+m-1}>0$, 则 u_j 这一项称为 m 领项. 因此, 引理阐明了所有 m 领项的 u_j(对任意 m, $j\leqslant j+m-1\leqslant n$)的和是非负的.

375

证明 对所有的 j(考虑 $k=j$), 我们有 $v_j\geqslant 0$. 对 $j=0, 1, \cdots, n-1$, 有 $v_j=(u_{j+1}+v_{j+1})^+$, 如下所示: $v_j=\max(0, u_{j+1}, u_{j+1}+u_{j+2}, \cdots, u_{j+1}+\cdots+u_n)$, 同理, $v_{j+1}=\max(0, u_{j+2}, \cdots, u_{j+2}+\cdots+u_n)$, 如果 $j=n-1$, $v_n=0$, 因此, $v_j=\max(0, u_{j+1}+\max(0, u_{j+2}, \cdots, u_{j+2}+\cdots+u_n))=\max(0, u_{j+1}+v_{j+1})$. 如果 $v_j=0\leqslant v_{j+1}+0$ 或 $0<v_j=v(j, n)=v_{j+1}+u_{j+1}$, 则

$$v_j\leqslant v_{j+1}+u_{j+1}1_{\{v(j,n)>0\}},$$

对 j 求和, 从 v_j-v_{j+1} 加到 v_0-v_n 得到引理中的和 $v_0-v_n=v_0\geqslant 0$. □

10.7.3 引理(极大不等式) 对超加性序列 $\{f_n\}$, 对所有的 n 和 x, $f_n(x)\geqslant 0$, 令 g 是任意非负可积函数集. 设

$$v:=\sup_{n\geqslant 0}(f_n-\sum_{0\leqslant j<n}g\circ T^j).$$

那么

$$E_I(g1_{\{v>0\}})\leqslant\sup_{n\geqslant 1}E_I(f_n)/n.$$

证明 对每个 x, 对序列 $u_i:=f_i-f_{i-1}-g\circ T^{i-1}$ 运用 Riesz 引理, 那么有

$$v(j,n)=\max_{k:j\leqslant k\leqslant n}(f_k-f_j-\sum_{i=j}^{k-1}g\circ T^i),\quad 0\leqslant j<n.$$

由 Riesz 引理可得,

$$\sum_{j=0}^{n-1}(f_{j+1}-f_j-(g\circ T^j))1_{\{v(j,n)>0\}}\geqslant 0,\quad n=1,2,\cdots$$

对每个 j, 因为 $f_1\geqslant 0$, 可得 $f_{j+1}\geqslant f_j+(f_1\circ T)\geqslant f_j$, 因此序列 $\{f_j\}$ 是非减的. 因而有

376

$$f_n=f_n-f_0=\sum_{j=0}^{n-1}f_{j+1}-f_j\geqslant\sum_{j=0}^{n-1}(f_{j+1}-f_j)1_{\{v(j,n)>0\}}\geqslant\sum_{j=0}^{n-1}(g\circ T^j)1_{\{v(j,n)>0\}}.$$

对 $j\leqslant k$, 有 $f_k-f_j\geqslant f_{k-j}\circ T^j$. 因而 $v(j, n)\geqslant v(0, n-j)\circ T^j$. 因为 $g\geqslant 0$, 所以

$$f_n\geqslant\sum_{j=0}^{n-1}(g1_{\{v(0,n-j)>0\}})\circ T^j,$$

$$E_I(f_n)\geqslant\sum_{k=1}^{n}E_I(g1_{\{v(0,k)>0\}}).$$

在引理 10.7.3 中, 当 $k\to\infty$ 时, 事件 $\{v(0, k)>0\}$ 递增到事件 $\{v>0\}$, 因而当 $n\to\infty$ 时,

$$\frac{1}{n}\sum_{k=1}^{n}1_{\{v(0,k)>0\}}\uparrow 1_{\{v>0\}}.$$

那么由条件期望的单调收敛定理(10.1.7)和关于 n 的上确界可以得到极大不等式. □

令 $\bar{f}:=\limsup_{n\to\infty}f_n/n$，关于 $\bar{f}$ 将证明下面的引理.

10.7.4 引理 我们有 $\bar{f}\geqslant\bar{f}\circ T$ a.s.，且对任意使得 $h\geqslant h\circ T$ a.e. 成立的可测函数 h，其中 T 为有限测度空间 $(X,\mathcal{A},\mu)$ 上的保测变换，我们有 $h\circ T=h$ a.e..

证明 首先，对所有的 n，$f_{n+1}\geqslant f_1+f_n\circ T$，因此，

$$\frac{f_{n+1}}{n+1}\geqslant\frac{f_1}{n+1}+\frac{f_n\circ T}{n+1}.$$

当 $n\to\infty$ 时，$f_1/(n+1)\to 0$，$n/(n+1)\to 1$，因此，取 lim sup 得到 $\bar{f}\geqslant\bar{f}\circ T$. 对第二部分，如果结论不成立，则存在有理数 q，使得 $\mu(h\circ T<q<h)>0$. 那么因为 μ 是有限的，$\mu(h\circ T<q)>\mu(h<q)$，但是因为 T 是保测的，所以这些测度相等. 故得出矛盾，引理 10.7.4 得证. □

现在继续定理 10.7.1 的证明，因为常数列 $\{n\}_{n\geqslant 1}$ 是可加的，序列 $\{f_n+n\}_{n\geqslant 1}$ 是超加的. 如果在定理 10.7.1 中，如同 γ 是对 $\{f_n\}$ 定义的一样，γ' 是对 $\{f_n+n\}$ 定义的函数，那么 $\gamma'\equiv\gamma+1$，$(f_n+n)/n=(f_n/n)+1$，定理对 $\{f_n+n\}$ 的叙述和对 $\{f_n\}$ 的叙述是等价的. 因此可假设对所有的 n，$f_n\geqslant n$. 那么 $f\geqslant 1$.

对 $r=1,2,\cdots$，令 $g_r:=\min(r,\bar{f}-1/r)$. 那么 g_r 是一个不变函数($g_r\circ T=g_r$ a.s.)，且是非负可积的. 对 $g=g_r$ 应用极大不等式(10.7.3). 因为处处有 $g_r<\bar{f}$，所以在引理 10.7.3 中处处有 $v>0$，因此，如定理(10.7.1)所定义的那样可以得到 [377]

$$g_r=E_Ig_r\leqslant\sup_n E_I(f_n)/n=\gamma.$$

令 $r\to\infty$ 得出

$$\bar{f}\leqslant\gamma\quad\text{a.s.}\tag{*}$$

为了证明 $\liminf f_n/n\geqslant\gamma$ a.s.，首先假设 $\{f_n\}$ 是可加的，令 $h:=f_1$，使得

$$f_n=\sum_{j=0}^{n-1}h\circ T^j.$$

任取 a，$0<a<\infty$. 注意

$$h\geqslant\min(h,a)=a-(a-h)^+,\tag{**}$$

因此

$$\liminf_{n\to\infty}f_n/n\geqslant a-\limsup_{n\to\infty}n^{-1}\sum_{0\leqslant j<n}(a-h)^+\circ T^j$$

对 $(a-h)^+$ 用(*)，那么注意对任意可加序列，$\gamma=E_I(h)$，和证明开始时一样，由(**)可得

$$\liminf_{n\to\infty}f_n/n\geqslant a-E_I((a-h)^+)=E_I(\min(h,a))$$

令 $a\uparrow+\infty$，那么对可加序列得到，

$$f_n/n\to\gamma\text{ a.s.}\tag{***}$$

在一般情况下，对每个 j，由 $f_k\circ T^j\leqslant f_{k+j}-f_j$ 可得出，

$$\sum_{j=0}^{n-1}f_k\circ T^j\leqslant\sum_{j=0}^{n-1}(f_{k+j}-f_j)\leqslant kf_{n+k-1},\tag{****}$$

因为 $f_1, \cdots, f_{k-1}$ 是非负非减的，因此对 $r=n, \cdots, n+k-1$，$f_r \leqslant f_{n+k-1}$. 对固定的 k，当 $n\to\infty$ 时，$(n+k-1)/n\to 1$，因此有

$$\liminf_{n\to\infty} f_n/n = \liminf_{n\to\infty} f_{n+k-1}/(n+k-1) = \liminf_{n\to\infty} f_{n+k-1}/n$$

$$\geqslant \frac{1}{k}\liminf_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f_k\circ T^j \quad 由(****)$$

$$=\frac{1}{k}E_I(f_k) \quad 由(***),用 f_k 替换 h.$$

378 取 k 的上确界，那么得到 $\gamma\leqslant\liminf_{n\to\infty} f_n/n$，因此由(*)可得，$f_n/n\to\gamma$ a. s..

现在令 $z_n := f_n/n$. 那么有 $z_n\to\sup_m E_I z_m=\gamma$，且对所有的 m，$E\gamma\geqslant EE_I z_m = Ez_m$. 由法图引理(4.3.3)得，$E\gamma\leqslant\liminf_{n\to\infty} Ez_n$，所以 $Ez_n\to E\gamma$. 因此，γ 是可积的当且仅当 $\sup_n Ez_n<\infty$，对于 L^1 收敛性，$0\leqslant(\gamma-z_n)^+\leqslant\gamma$，所以如果 $E\gamma<\infty$ 那么由控制收敛(定理 4.3.5)可得，$E(\gamma-z_n)^+\to 0$. 因为 $E(\gamma-z_n)\to 0$，所以有 $E(\gamma-z_n)^-\to 0$ 和 $E|\gamma-z_n|\to 0$，故定理 10.7.1 证毕. □

例(随机积)：假设 M 是一个半群(通常是矩阵或线性算子)，使得结合积运算有定义. 假设 $\|\cdot\|$ 是从 M 映射到正实数的*子乘法函数*(submultiplicative function)，即对所有的 $A, B\in M$，$\|AB\|\leqslant\|A\|\|B\|$. 令 A_1 是从一个概率空间映射到 M 的可测函数(对于 M 上的 σ-代数，使得积是联合可测的且 $\|\cdot\|$ 是可测的)，令 T 是概率空间上的保测变换，对 $n=1, 2, \cdots$ 令 $A_n := A_1\circ T^{n-1}$，$f_n := \log\|A_n A_{n-1}\cdots A_1\|$，那么 $\{f_n\}$ 是超加的.

有一种情况，令 M 是 $\mathbb{R}^k$ 映射到自身的所有线性变换的集合，它可由 $k\times k$ 的矩阵表示，如果 $\|\cdot\|$ 是 $\mathbb{R}^k$ 上通常的范数，且 $A\in M$，令 $\|A\|$ 为范数算子 $\|A\| := \sup\{|Ax|/|x|: x\neq 0, x\in\mathbb{R}^k\}$，容易看出 $\|\cdot\|$ 是子乘法的.

如果 A_j 是 M 的独立同分布元素，则不改变它们的联合分布，可假设它们是关于 M 的可数笛卡儿积的坐标，且有积概率分布和 σ-代数，这里 T 是移位变换 $\{A_j\}\mapsto\{A_{j+1}\}$. 那么由科尔莫戈罗夫 0-1 律(8.4.4)知，不变 σ-代数是平凡的，所以 f_n/n 几乎必然收敛到某个常数. 然而对计算常数值它可能是非平凡的. (参考 Furstenberg 和 Kesten, 1960; Furtenberg, 1963; Ledrappier, 1984).

例(格指标变量)：超加序列的理论要追溯到逾渗理论(Hammersley 和 Welsh, 1965). 例如，设一平面被边长为 1 的方格分解. 假设对格中由两个方格的公共边形成的每条直线段 S，存在一个给定非负随机变量 $X(S)$. 对格的任意两个竖边 u 和 v，令 $L(u, v)$ 是所有 $X(S)$(连接 u 和 v 的所有序列段 S)和的下确界. 那么对任意竖边 u，v 及 w，$L(u, w)\leqslant L(u, v)+L(v, w)$. 因此，$L$ 是由格中的点构成的次加性过程. 限定到垂线 $(n, 0)$，其中 $n=1, 2, \cdots$，可得到如上所定义的次加性

379 过程.

习题

1. 设 $X_1, X_2, \cdots$ 是独立同分布实值随机变量且 $E|X_1|<\infty$，$S_n := X_1+\cdots+X_n$. 证明：$|S_n|$ 是一个次加性序列，并通过对 X_1 的积分求出 $|S_n|/n$ 的极限.
2. 设 $(X, \mathcal{A}, P)$ 是一个概率空间，T 是 X 映上到其自身的保测变换. Y 是 X 上的实值随机变量，且对 $j=1, 2, \cdots$，$Y_j := Y\circ T^j$. 令 $S_0 := 0$，$S_n := Y_1+\cdots+Y_n$，$f_n := \max\{S_k - S_j: 0\leqslant j\leqslant k\leqslant n\}$. 证明：$\{f_n\}$ 是次加的.
3. 设 $(X, \|\cdot\|)$ 是一个线性赋范空间. 如果从 X 映射到其自身的线性函数 A 满足 $\|A\| := \sup\{\|Ax\|: \|x\|=1\}<\infty$，那么 A 称为有界算子，$\|A\|$ 称为 A 的算子范数.

(a)证明：算子范数在所有的有界线性算子集合上是子乘法的.

(b)在具有通常的欧几里得范数的$\mathbb{R}^2$上，设A_1，A_2，…是独立同分布线性算子，每个都可由下面的矩阵来表示，取每个矩阵的概率为1/3：

$$\begin{bmatrix}1&0\\0&1\end{bmatrix},\begin{bmatrix}2&0\\0&2\end{bmatrix},\begin{bmatrix}1&1\\0&1\end{bmatrix}.$$

求在随机积的例子中的极限常数(本节的末尾).

4. 设$\{f_n\}$是一个次加性序列且$A:=\{\sup_n f_n>0\}$. 证明：$\int_A f_1 dP \geqslant 0$.［提示：参考极大遍历引理(8.4.2).］

5. 次加性遍历定理通常叙述为：对随机变量列$\{f_{m,n}\}_{0\leqslant m<n}$，其联合分布和$\{f_{m+1,n+1}\}_{0\leqslant m<n}$的联合分布是相同的，当$k<m<n$时，它满足$f_{k,n}\leqslant f_{k,m}+f_{m,n}$. 证明：如果$\{f_n\}$和$T$满足定理10.7.1的假设条件，那么$f_{m,n}:=f_{n-m}\circ T^m$满足上述次加性遍历定理.

6. 相反地，假设$\{f_{m,n}\}_{0\leqslant m<n}$满足习题5的条件. 令$\Omega$是笛卡儿积$\prod_{0\leqslant m<n}\mathbb{R}_{m,n}$，坐标为$\{x_{m,n}\}_{0\leqslant m<n}$，令$P$是$\Omega$上$\{f_{m,n}\}_{0\leqslant m<n}$的法则. 证明：$T$：$\{x_{m,n}\}\mapsto\{x_{m+1,n+1}\}$是$\Omega$上的保测变换，$f_n:=x_{0,n}$是一个次加性序列，且如习题5有$x_{m,n}=f_{n-m}\circ T^m$.

7. 在T. Liggett(1985)的定理中，通过假定次可加性$f_{n,k}\leqslant f_{k,m}+f_{m,n}(0=k<m<n)$可以减弱习题5的假设条件，且联合分布是相等的： [380]

(a)对每个$m\geqslant 0$，$\{f_{m+1,m+k+1}\}_{k\geqslant 1}$和$\{f_{m,m+k}\}_{k\geqslant 1}$有相同的联合分布.

(b)对每个$k\geqslant 1$，$\{f_{nk,(n+1)k}\}_{n\geqslant 1}$和$\{f_{(n+1)k,(n+2)k}\}_{n\geqslant 1}$有相同的联合分布.

证明：习题5的假设可推出(a)和(b).

8. 证明：(a)对每个m和n，且$0\leqslant m<n$，$f_{m+1,n+1}$和$f_{m,n}$有相同的分布.

9. 证明：(a)和(b)不能推出$\{f_{m+1,n+1}\}_{0\leqslant m<n}$和$\{f_{m,n}\}_{0\leqslant m<n}$有相同的联合分布.［提示：令$f_{0,n}$是独立同分布的且法则为$N(0,1)$，对所有的$m$和$k$，除了$f_{1,3}=-f_{0,2}$，令$f_{m,m+k}=f_{0,k}$.］

注释

10.1节 对概率空间$(\Omega,\mathcal{B},P)$，对Ω上满足$E|X|<\infty$的随机变量X和Y，Kolmogorov (1933，Chap，V，Sec.4)定义了X关于Y的条件期望$E(X|Y)$. 这与$E(X|\mathcal{B}_Y)$是相同的，其中$\mathcal{B}_Y$是Y在其上可测的最小σ-代数. 同理，当$Y:=(Y_1,\cdots,Y_n)$是随机向量或n维随机变量时，$E(X|Y):=E(X|Y_1,\cdots,Y_n)=E(X|\mathcal{B}_Y)$. 令$(S,\mathcal{T})$是可测空间，$Y$是由$\Omega$映射到$S$的可测函数. Halmos(1950，p.209)定义了关于Y的条件期望. 然而关于函数的条件期望可能比关于σ-代数的条件期望更特殊，因为Y在其上可测的最小σ-代数总是可以取到的，通过令$S=\Omega$，Y为单位元，$\mathcal{T}$为$\mathcal{B}$的子σ-代数，也可以得到另一方向的条件期望. Doob(1953，p.17—18，623)定义了关于子σ-代数的条件期望. Kolmogorov(1933)用Radon-Nikodym定理给出了存在性的证明，现在已被广泛应用.

Black well和Dubins(1963)证明了条件期望控制收敛定理(10.1.8)的逆：如果$f_n\geqslant 0$且$f_n\to f_0$ a.s.，其中对所有的$n\geqslant 0$，$f_n\in\mathcal{L}^1(P)$，且如果$E\sup_n f_n=+\infty$，那么在某概率空间上存在随机变量$\{g_n\}_{n\geqslant 0}$，它和$\{f_n\}_{n\geqslant 0}$有相同的分布，且可测集的σ-代数$\mathcal{C}$满足$P(E(g_n|\mathcal{C})\to E(g_0|\mathcal{C}))=0$.

10.2节 Doob(1938)和Halmos(1941)首次着手证明正则条件概率(或者更一般地称为商测度)的存在性. 他们的叙述的条件是不充分的，事实上，在定理10.2.2中，要求可分性，而不要求完备性. Andersen and Jensen(1948)给出了关于无穷积空间测度存在性叙述的一个反例. Dieudonné (1948，p.39—51)证明了正则条件概率不是在任意可分度量空间上存在而在紧度量空间上是存在的. 这个事实和Ulam定理的完全可分度量空间的结果相差不远了.

Hölder(1889)证明了詹森不等式对$\mathbb{R}^1$的情况，如果X是简单函数且f处处有一阶导数，然后Jensen(1906，p.186)对任意连续凸函数给出了证明．Hölder的名字命名了另一个著名的不等式(5.1.2)，因此，我对(10.2.6)不建议用Hölder-Jensen不等式来命名．正则条件概率的存在性的证明和它们在证明条件詹森不等式的应用是基于Bauer(1968. §10.3)的．

381

10.3 节 10.3.4 前面的例子有一个逆：对概率空间$(\Omega, \mathcal{B}, P)$上的任意非负下鞅序列$\{X_n, \mathcal{B}_n\}$，存在一个概率空间$(V, \mathcal{C}, Q)$和鞅序列$\{Y_n, \mathcal{C}_n\}$，使得在$\mathbb{R}^\infty$上，$\{X_n\}_{n\geq1}$和$\{Y_n\}_{n\geq1}$的分布是相同的．对非负下鞅$\{X_t\}_{0\leq t<\infty}$有类似的结果(Gilat，1977)．

Random House(1987)给出了鞅的四个含义，两个关于马的挽具，一个是关于航船的船帆，一个是关于赌局，这才引出了它的数学含义．某些意义涉及很多分支．语源学追溯到1600年前中世纪的法国，那时有两个可能的分支：一个是通过martegalo——Martigues镇女居民的普洛旺斯语，另一个通过西班牙语almártaga——挽具，阿拉伯语almarta'ah——脉络．

鞅的数学理论大约在1940年开始蓬勃发展．Ville(1939，p.111—130)对随机变量序列开始用鞅这个词．Doob(1940)开始了严格深入的理论研究．

Vitali(1904—1905，1907)首次给在$\mathbb{R}$中的有界测度集上关于勒贝格测度定理给出了10.3.5和10.3.6的形式．

10.4～10.5 节 Doob(1940)首次证明了这两节中关于鞅的大部分结果(他用"ε性质"而不是"鞅")．早期Lévy(1937)证明了完备鞅的几乎必然收敛性，使得X_∞是某事件的示性函数．Doob(1953，p.294)介绍并解决了下鞅(并且由对称性可得上鞅)的情况，他那时称为半鞅．Doob(1953，p.629—634)给出了鞅的注，提到了Ville(1939)，且主要描述了Doob自己的研究与Andersen and Jessen(1946，1948b)的关系．由引理10.5.6得10.5节的证明与Lamb(1973)中相同．大多数其他的证明计算了任意区间$[a, b]$上穿(upcrossing)的次数，$X_{n(2j-1)}(\omega)\leq a<b\leq X_{n(2j)}(\omega)$，$j=1, 2, \cdots$，$n(1)<n(2)<\cdots$，Lamb的证明比较简短．Baez-Duarte(1968)和P-A. Meyer(1972)不用上穿给出了另一种证明．更深入的探讨见Hall and Heyde(1980)．

10.6 节 Doob(1940，p.458；1953)首次讨论了逆鞅．本节结果的不同证明在Doob(1953，p.329)给出．在那本书中Doob首次介绍了逆下鞅(和下鞅一样)．

10.7 节 次加性遍历定理(10.7.1)归功于Kingman(1968)．证明由Neveu(1983)给出．大多数的应用是用次加性而不是超加性．Riesz(1945，p.225)证明了他的m-领项引理(10.7.2)．

次加性遍历定理之前的证明把经典的Birkhoff遍历定理(8.4.1)应用到可加序列．Neveu的证明没有，因而他同时证明了Birkhoff定理．Kingman(1973，1976)研究了次加性过程．

Liggett(1985)在弱的条件(在习题7提到)下证明了次加性遍历定理的结论，暗示了这个扩张定理的广泛应用．Levental(1988)给出了Birkhoff定理和Liggett定理的新的证明．

382

参考文献

带星号的资料在次级来源中被讨论过，在原著中并没见到．

Andersen, Erik Sparre, and Børge Jessen (1946). Some limit theorems on integrals in an abstract set. *Danske. Vid. Selsk. Mat.-Fys. Medd. 22*, no. 14. 29 pp.

———, and ——— (1948a). On the introduction of measures in infinite product sets. *Danske Vid. Selsk. Mat.-Fys. Medd. 25*, no. 4. 8 pp.

———, and ——— (1948b). Some limit theorems on set-functions. *Danske Vid. Selsk. Mat.-Fys. Medd. 25*, no. 5. 8 pp.

Baez-Duarte, Luis (1968). Another look at the martingale theorem. *J. Mathematical Analysis and Appls. 23*: 551–557.

Bauer, Heinz (1968). *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie*. De Gruyter, Berlin. Publ. in English as *Probability Theory and Elements of Measure Theory*. Transl. Lisa Rosenblatt. Holt, Rinehart & Winston, New York, 1972.

Blackwell, David, and Lester E. Dubins (1963). A converse to the dominated convergence theorem. *Illinois J. Math. 7*: 508–514.

Dieudonné, Jean (1948). Sur le théorème de Lebesgue-Nikodym (III). *Ann. Inst. Fourier* (Grenoble) (Ser. 2) *23*: 25–53.

Doob, Joseph L. (1938). Stochastic processes with an integral-valued parameter. *Trans. Amer. Math. Soc. 44*: 87–150.

——— (1940). Regularity properties of certain families of chance variables. *Trans. Amer. Math. Soc. 47*: 455–486.

——— (1953). *Stochastic Processes*. Wiley, New York.

Furstenberg, Harry (1963). Non-commuting random products. *Trans. Amer. Math. Soc. 108*: 377–428.

———, and Harry Kesten (1960). Products of random matrices. *Ann. Math. Statist. 31*: 457–469.

Gilat, David (1977). Every nonnegative submartingale is the absolute value of a martingale. *Ann. Probability 5*: 475–481.

Hall, Peter, and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

Halmos, Paul R. (1941). The decomposition of measures. *Duke Math. J. 8*: 386–392.

——— (1950). *Measure Theory*. Van Nostrand, Princeton.

Hammersley, J. M., and D. J. A. Welsh (1965). First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. *Bernoulli-Bayes-Laplace Anniversary Volume*, ed. J. Neyman and L. M. Le Cam, pp. 61–110. Springer, New York.

Hölder, Otto (1889). Ueber einen Mittelwerthssatz. *Gesellschaft der Wissenschaften, Göttingen, Nachrichten* (= *Göttinger Nachrichten*), pp. 38–47.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalites entre les valeurs moyennes. *Acta Math. 30*: 175–193.

Kingman, John F. C. (1968). The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc.* (Ser. B) *30*: 499–510.

——— (1973). Subadditive ergodic theory (With discussion). *Ann. Probability 1*: 883–899.

——— (1976). Subadditive processes. *École d'été de probabilités de St-Flour*, 1975. *Lecture Notes in Math.* (Springer) *539*: 167–223.

Kolmogoroff, Andrei N. [Kolmogorov, A.N.] (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin. 2d English ed., *Foundations of the Theory of Probability*, transl. ed. Nathan Morrison. Chelsea, New York, 1956.

Lamb, Charles W. (1973). A short proof of the martingale convergence theorem. *Proc. Amer. Math. Soc. 38*: 215–217.

Ledrappier, F. (1984). Quelques propriétés des exposants caractéristiques. *École d'été de probabilités de St-Flour*, 1982. *Lecture Notes in Math.* (Springer) *1097*: 305–396.

Levental, Shlomo (1988). A proof of Liggett's version of the subadditive ergodic theorem. *Proc. Amer. Math. Soc. 102*: 169–173.

Lévy, Paul (1937). *Théorie de l'addition des variables aléatoires.* 2d. ed. 1954. Gauthier-Villars, Paris.

Liggett, Thomas M. (1985). An improved subadditive ergodic theorem. *Ann. Probability 13*: 1279–1285.

Meyer, Paul-André (1972). *Martingales and Stochastic Integrals I. Lecture Notes in Math. 284* (Springer): Berlin.

Neveu, J. (1983). Courte démonstration du théorème ergodique sur-additif. *Ann. Inst. Henri Poincaré* (Ser. B) *19*: 87–90.

Random House Dictionary of the English Language (1987), 2d ed., unabridged, Random House, New York.

Riesz, F. (1945). Sur la théorie ergodique. *Comment. Math. Helvetici 17*: 221–239.

*Ville, Jean (1939). *Étude critique de la notion de collectif.* Gauthier-Villars, Paris.

*Vitali, Giuseppe (1904–05). Sulle funzioni integrali. *Atti Accad. Sci. Torino 40*: 1021–1034.

——— (1907). Sull'integrazione per serie. *Rend. Circolo Mat. Palermo 23*: 137–155.



第 11 章　可分度量空间上的依 L 收敛

迄今为止，我们运用中心极限定理(9.3～9.5 节)主要研究了在有限维欧几里得空间$\mathbb{R}^k$上的依 L 收敛．现在将在更一般(可能无穷维)空间上研究依 L 收敛，下面是一些例子．

在某一概率空间$(\Omega, \mathcal{B}, P)$上定义$x(t, \omega)(\omega \in \Omega)$表示一个随机运动的质点在时刻$t$的位置．对于每个$\omega$，定义一个连续函数$t \mapsto x(t, \omega)$，其中$t \in [0, 1]$，$x$是实值函数(质点沿着一条直线运动，或者只考虑质点位置的一个坐标)．这样$x(\cdot, \omega) \in C[0, 1]$，其中$C[0, 1]$是定义在$[0, 1]$上的连续实值函数的全体．在空间$C[0, 1]$上有一个常用的范数：上确界范数$\|f\| := \sup\{|f(t)| : 0 \leqslant t \leqslant 1\}$，这样$C[0, 1]$空间对于通常定义为$d(f, g) := \|f - g\|$的距离$d$是一个完备可分度量空间．这可以很好地去逼近过程$x$，例如，通过一个过程$y_n$，使得对每个$k = 1, 2, \cdots, n$，$y_n(\cdot, w)$在区间$[(k-1)/n, k/n]$上是线性的．因此，这可以帮助定义$y_n$在$C[0, 1]$上依 L(或者依概率或者几乎处处)收敛到$x$.

11.1　法则和收敛性

首先回忆一下 7.1 节和 9.3 节的一些定义．设$(S, \mathcal{T})$是一个拓扑空间，由$\mathcal{T}$生成的σ-代数(包含$\mathcal{T}$的最小σ-代数)$\mathcal{B}$称为博雷尔σ-代数，$\mathcal{B}$中的集合称为博雷尔集，定义在$\mathcal{B}$上的测度称为博雷尔测度．一个 L 是博雷尔概率测度，换句话说，它是定义在$\mathcal{B}$上的一个概率测度．L 的最简单的例子是点质量δ_x，其中$\delta_x(A) := 1_A(x)$，$\forall A \in \mathcal{B}$.

S或$(S, \mathcal{T})$称为波兰空间，当且仅当$\mathcal{T}$对使得S是完备可分的度量d是可度量的．回忆一个可分度量空间(S, e)是波兰空间，当且仅当(S, e)是其完备化空间中开集的可数交(定理 2.5.4).

令$C_b(S) := C_b(S, \mathcal{T})$是由所有$S$上的实值有界连续函数组成的空间．回忆 L 的序列$P_n$依 L 收敛于$P$，记为$P_n \to P$，当且仅当$\int f \mathrm{d}P_n \to \int f \mathrm{d}P$对于任意$f \in C_b(S)$成立(见 9.3 节)．注意到，对于一个度量空间$(S, d)$，这种收敛仅依赖于$S$的拓扑，与特殊的度量没有关系．

回忆如果一个取值于可分度量空间上的随机变量列几乎必然或依概率收敛，那么它们就依 L 收敛(命题 9.3.5)．另一方面，法则序列δ_n(n处的单位质量)在$\mathbb{R}$上不收敛(可以考虑$f(t) := \cos(\pi t)$).

回忆(2.1 节)对于任意拓扑空间$(S, \mathcal{T})$和任意的集合$A \subset S$，记$\bar{A}$为A的闭包，int A为A内部(包含在A中的最大开集)，∂A为A的边界，$\partial A := \bar{A} \setminus \text{int } A$.(注意$\partial A$总是闭的.)对于定义在$\mathcal{B}$上的测度$\mu$，集合$A$称为连续集当且仅当$\mu(\partial A) = 0$.

例如，$\mathbb{R}$上的区间$[a, b]$是一个点质量δ_x的连续集，当且仅当x不等于a或b. 下面给出度量空间中依 L 收敛的几个等价条件，这被称为 portmantéau 定理．

11.1.1 定理　对于度量空间(S, d)中的法则P_n和P，下列命题等价：

(a)$P_n \to P$.

(b)对所有的开集U，$\lim\limits_{n\to\infty} \inf P_n(U) \geqslant P(U)$.

(c)对所有的闭集F，$\lim\limits_{n\to\infty} \sup P_n(F) \leqslant P(F)$.

(d)对 P 中的所有连续集 A，$\lim_{n\to\infty}P_n(A)=P(A)$.

例：令 P_n 为单位质量 $\delta_{1/n}$，且在$\mathbb{R}$ 中 $P=\delta_0$. 那么 $P_n\to P$. 在(b)中，设 U 是开的半射线$(0,\infty)$，并且在(c)中，令 F 为闭的半射线$(-\infty,0]$. 那么即使(b)中的 lim inf 和(c)中的 lim sup 就是实际的极限($P_n\equiv 1$ 和 $P_n(F)\equiv 0$)，(b)和(c)中的不等式也是严格的．F 和 U 都不是 P 中的连续集．这个例子可以帮助回忆并且直观地看到 portmanteau 定理的不同条件．

证明 (a)⇒(b)：令 U 是开的，$F:=S\setminus U$，$\varepsilon>0$，$d(x,F):=\inf_{y\in F}d(x,y)$(如在 2.5.3 和定理

386

7.1.1 中那样)．令 $f_m(x):=\min(1,md(x,F))$，因此 $0\leqslant f_m\uparrow 1_U$(如引理 9.3.2 的证明中那样)．令 $F_m:=f_m^{-1}(\{1\})$. 那么 m 充分大时，有

$$\liminf_{n\to\infty}P_n(U)\geqslant\lim_{n\to\infty}\int f_m\mathrm{d}P_n=\int f_m\mathrm{d}P\geqslant P(F_m)>P(U)-\varepsilon,$$

由此可以推出(b).

显然，通过取其补集(b)和(c)是等价的．如果(b)和(c)成立，那么对任意博雷尔集 A，

$$\begin{aligned}P(\operatorname{int}A)&\leqslant\liminf_{n\to\infty}P_n(\operatorname{int}A)\leqslant\liminf_{n\to\infty}P_n(A)\leqslant\limsup_{n\to\infty}P_n(A)\\&\leqslant\limsup_{n\to\infty}P_n(\bar{A})\leqslant P(\bar{A}).\end{aligned}$$

如果 A 是 P 中的连续集，那么 $P(\operatorname{int}A)=P(\bar{A})$，因此 $P_n(A)\to P(A)$，即证得(d).

现在假设(d)成立．令 $f\in C_b(S)$. 对于不同的 y 值，集合 $F_y:=\{x\in S: f(x)=y\}$ 是不相交的，因此对于至多可数多个 y 的值，$P(F_y)>0$. 加一个常量到 f 上，不影响 $\int f\mathrm{d}P_n$ 收敛到 $\int f\mathrm{d}P$. 我们可以假设 $P(F_0)=0$. 那么对于 $\varepsilon>0$ 和每个整数 k，令

$$B_{k,\varepsilon}:=\{x:k\varepsilon\leqslant f(x)<(k+1)\varepsilon\}.$$

那么$\partial B_{k,\varepsilon}\subset F_{k\varepsilon}\cup F_{(k+1)\varepsilon}$. 所以存在任意小的 ε，使得对于所有的整数 k，$B_{k,\varepsilon}$是 P 的连续集．那么对于所有的 k，当 $n\to\infty$ 时，

$$(P_n-P)(B_{k,\varepsilon})\to 0,$$

由于 f 是有界的，对于给定的 ε，只有有穷多个 k，使得 $B_{k,\varepsilon}$是非空的．这时有

$$\begin{aligned}\int f\mathrm{d}P-\varepsilon&\leqslant\sum_k k\varepsilon P(B_{k,\varepsilon})=\lim_{n\to\infty}\sum_k k\varepsilon P_n(B_{k,\varepsilon})\leqslant\liminf_{n\to\infty}\int f\mathrm{d}P_n\\&\leqslant\limsup_{n\to\infty}\int f\mathrm{d}P_n\leqslant\lim_{n\to\infty}\sum_k(k+1)\varepsilon P_n(B_{k,\varepsilon})\\&=\sum_k(k+1)\varepsilon P(B_{k,\varepsilon})\leqslant\int f\mathrm{d}P+\varepsilon.\end{aligned}$$

令 $\varepsilon\downarrow 0$ 即可证得． □

在定理 11.1.1 后面提到的例子 $\delta_{1/n}\to\delta_0$ 中，可以得出当 $t\neq 0$ 时，分布函数 $F_n(t)$ 收敛到 $F(t)$，其中 $F(t)$是不连续的．这表明了下面的事实．

11.1.2 Helly-Bray 定理 如果 P_n 和 P 是$\mathbb{R}$ 上的法则，其分布函数分别是 $F_n(n=1,2,\cdots)$和 F，那

387

么 $P_n\to P$ 当且仅当对所有的 t，$F_n(t)\to F(t)$，其中 F 是连续的．

证明 必要性：F 在 t 处是连续的，当且仅当$(-\infty,t]$是 P 的连续集，那么根据 portmanteau 定理(a)⇒(d)，可以证得必要性，这在前面也被直接证明了(定理 9.3.6).

充分性：令 U 是$\mathbb{R}$ 中的任意开集，$\varepsilon>0$，那么 U 是开区间的可数并(具有有理端点的开区间形成了通常拓扑的基)，因此，存在有限并 $V=\bigcup_{1\leqslant j\leqslant m}(a_j,b_j)\subset U$，并且 $P(V)>P(U)-\varepsilon$. 这里选择的

a_j 和 b_j 满足对所有的 j，$[a_j, b_j] \subset U$. 那么取 a_j 较小，b_j 较大，可以假设这些点对于 F 或 F_n 是连续的，因为不连续点集是可数集. 那么有

$$\liminf_{n\to\infty} P_n(U) \geqslant \lim_{n\to\infty} P_n(V) = P(V) > P(U) - \varepsilon,$$

则根据 portmanteau 定理中(b)⇒(a)这个事实即可证得充分性. □

回忆在可分度量空间中，依概率收敛可以推出依 L 收敛(命题 9.3.5)，但是反之一般是不成立的. 在下面的情况下，依 L 收敛可以推出依概率收敛.

11.1.3 命题 如果(S, d)是一个度量空间，$p \in S$，X_n 是概率空间$(\Omega, \mathcal{B}, P)$上的随机变量，且 X_n 满足$\mathcal{L}(X_n) \to \delta_p$，那么 X_n 依概率收敛到 p.

证明 令 $f_k(x) := \max(0, 1 - kd(x, p))$，那么 $f_k \in C_b(S)$ 且对 $\forall k$，当 $n \to \infty$ 时，$P\{d(X_n, p) < 1/k\} \geqslant Ef_k(X_n) \to 1$. □

在命题 11.1.3 中，依概率收敛不能总是被几乎必然收敛所替代(见习题 1).

11.1.4 命题 对任意拓扑空间$(S, \mathcal{T})$和 S 的博雷尔集上的测度 μ，μ 的连续集形成一个代数.

证明 对任意集合 $A \in S$，$\partial A = \partial(S \setminus A)$，因此，$A$ 是连续集当且仅当 A 的补集是连续集. 空集 $\varnothing$ 和 S 总是连续集. 对任意集合 C，其闭包 $\bar{C}$ 可以记为$(C)^-$. 其次，对任意两个集合 A 和 B，$\mathrm{int}(A \cup B) \supset \mathrm{int}\, A \cup \mathrm{int}\, B$ 且 $(A \cup B)^- = \bar{A} \cup \bar{B}$，因此，

$$\partial(A \cup B) \subset (\bar{A} \cup \bar{B}) \setminus (\mathrm{int}\, A \cup \mathrm{int}\, B) \subset \partial A \cup \partial B.$$

所以，如果 A 和 B 是连续集，那么 $A \cup B$ 也是连续集. □ 388

例如，令 A 和 B 分别为$\mathbb{R}$上的有理数集和无理数集，那么根据上面的证明我们可以得出 $\mathrm{int}\, A \cup \mathrm{int}\, B \subset \mathrm{int}(A \cup B)$ 和 $\partial(A \cup B) \subset \partial A \cup \partial B$ 是严格成立的.

习题

1. 试举例：实值随机变量 X_n，$\mathcal{L}(X_n) \to \delta_0$，使得 X_n 不几乎必然收敛到 0.(根据命题 11.1.3，X_n 依概率收敛到 0).[提示：见 9.2 节习题 1.]
2. 证明：对任意具有离散拓扑的可数集 S，依测度收敛等价于依全变差收敛(在 9.3 节所定义的).
3. 试举例：P 的连续集不能形成一个 σ-代数.[提示：对于[0, 1]上的勒贝格测度，考虑单元素集.]
4. 如果 F_n 和 F 是$\mathbb{R}$上的分布函数，F 是连续的，并且对 $\forall t$，$F_n(t) \to F(t)$，证明：$F_n(t)$ 一致收敛到 $F(t)$，即当 $n \to \infty$ 时，$\sup_t |(F_n - F)(t)| \to 0$.[提示：给定 $\varepsilon > 0$，寻找一个有限集，使得 F 在其上的值在[0, 1]上是 ε 稠密的，然后运用 F_n 和 F 的单调性.]
5. 令 $x(n)$ 是$\mathbb{R}$上的任意点，满足当 $n \to \infty$ 时，$|x(n)| \to \infty$. 证明：法则 $\delta_{x(n)}$ 不收敛.
6. 试举例：在$\mathbb{R}$上，法则 $P_n \to P$，当 F 关于 t 连续时，分布函数 $F_n(t) \to F(t)$(根据 Helly-Bray 定理)，但是在一个稠密集中，对于 t，$F_n(t)$ 不收敛于 $F(t)$.[提示：定义$\mathbb{R}$上的一个法则 P，使得对任意一个有理数 q 都有 $P(\{q\}) > 0$. 令 $P_n(A) = P(A - 1/n)$，其中 A 为任意博雷尔集.]
7. 对$\mathbb{R}^2$上的一个法则 P，其分布函数定义为 $F(x, y) := P((-\infty, x] \times (-\infty, y])$，定义 P_n 依 L 收敛于 P，其分布函数分别为 F_n 和 F. 证明：对几乎所有的(x, y)，$F_n(x, y)$ 关于$\mathbb{R}^2$上的勒贝格测度 λ^2 收敛于 $F(x, y)$.
8. 证明：在习题 7 中，要么 $F_n(x, y)$ 收敛于 $F(x, y)$，$\forall (x, y) \in \mathbb{R}_2$，要么对于$(x, y)$的不可数集，$F_n(x, y)$ 不收敛于 $F(x, y)$.
9. 在习题 7 和 8 的条件下，假设对$\mathbb{R}^2$中的每个点(x, y)，有 $P(\{(x, y)\}) = 0$. 证明：对于某个坐标旋转 $x' = x\cos\theta - y\sin\theta$，$y' = x\sin\theta + y\cos\theta$，$P_n$ 在坐标(x', y')上的分布函数 $G_n(x', y')$ 收敛到 P 相应的分布函数.
10. 称一个拓扑空间中的集合 A 是完全的(full)当且仅当 $\mathrm{int}\, A = \mathrm{int}\, \bar{A}$. 证明或反证：在$\mathbb{R}$中，完全集形成一个代数.

389

11. 对$\mathbb{R}$上的任意非减函数 F，令 $F(t^-) := \lim_{x\uparrow t} F(x)$，$\forall t\in\mathbb{R}$．假设 F_n 和 F 是分布函数，满足 $F_n(t)\to F(t)$ 和 $F_n(t^-)\to F(t^-)$，$\forall t\in\mathbb{R}$．证明：$\lim_{n\to\infty}\sup_t|(F_n-F)(t)|=0$.（换句话说，$F_n$ 在$\mathbb{R}$上一致收敛到 F）.
[提示：F 连续时的情况是习题4. 在这种情况下，给定 $\varepsilon>0$，寻找一个有限集 G，使得 G 中的 $F(t)$ 对于 t 在 F 的值域中是 ε 稠密的，且对于任意的跳跃 $F(t)-F(t^-)\geqslant\varepsilon$ 有 $t\in G$. 然后证明如果对每个 $t\in G$，一个非减函数 H 在 t 和 t^- 上的值是在 F 的 ε 邻域内的，那么$\sup_t|(F-H)(t)|\leqslant 4\varepsilon$.]

11.2 利普希茨函数

在任意拓扑空间上都可以定义连续函数．在欧几里得空间或者更一般的可微流形上可以定义任意阶可导函数．在一个一般的度量空间中，光滑性的最自然的形式是如下的利普希茨条件.

令$(S,\ d)$是一个度量空间，回忆对于 S 上的实值函数f，利普希茨半范数定义为

$$\|f\|_L := \sup_{x\neq y}|f(x)-f(y)|/d(x,y),$$

（见6.1节），那么 $\|f\|_L=0$ 当且仅当f是一个常值函数．上确界范数为 $\|f\|_\infty := \sup_x|f(x)|$. 令

$$\|f\|_{BL} := \|f\|_L + \|f\|_\infty \text{ 和 } BL(S,d) := \{f\in\mathbb{R}^S : \|f\|_{BL}<\infty\}.$$

这里“BL”表示“有界的利普希茨”，$BL(S,\ d)$是 S 上所有实值有界利普希茨函数的集合.

例：如果$f(x)=\sin x$ 和 $g(x):=\cos x$，那么f和 g 是$\mathbb{R}$上的有界利普希茨函数．函数$f(x):=2x$ 和 $g(x):=|x|$ 是利普希茨函数，但不是有界的．函数$f(x):=x^2$ 不是$\mathbb{R}$上的利普希茨函数.

这里的 $\|\cdot\|_{BL}$不仅是一个范数，还具有子乘法的性质.

11.2.1 命题 *$BL(S,\ d)$是向量空间，$\|\cdot\|_{BL}$是其上的范数，对于$\forall f,\ g\in BL(S,\ d)$，$fg\in BL(S,\ d)$且 $\|fg\|_{BL}\leqslant\|f\|_{BL}\|g\|_{BL}$.*

证明 向量空间的性质是显然的．由于在 $BL(S,\ d)$中，$\|\cdot\|_L$是半范数，$\|\cdot\|_\infty$是范数，$\|\cdot\|_{BL}$是范数，显然，$\|fg\|_\infty\leqslant\|f\|_\infty\|g\|_\infty$. 对$\forall x,\ y\in S$,

$$\begin{aligned}|f(x)g(x)-f(y)g(y)| &\leqslant |f(x)(g(x)-g(y))|+|g(y)(f(x)-f(y))|\\ &\leqslant (\|f\|_\infty\|g\|_L+\|g\|_\infty\|f\|_L)d(x,y),\end{aligned}$$

[390] 所以$fg\in BL(S,\ d)$并且范数不等式成立. □

另外，考虑格运算$f\vee g:=\max(f,\ g)$，$f\wedge g:=\min(f,\ g)$. 像加法一样，这些运算也保持利普希茨性质.

11.2.2 命题 *对任意的f_1，…，f_n 和 $*=\vee$或$\wedge$，有*

(a) $\|f_1*\cdots*f_n\|_L\leqslant\max_{1\leqslant i\leqslant n}\|f_i\|_L$.

(b) $\|f_1*\cdots*f_n\|_{BL}\leqslant 2\max_{1\leqslant i\leqslant n}\|f_i\|_{BL}$.

证明 通过对 n 进行归纳来证明(a)，只需证明 $n=2$ 时成立即可．如果$(f\vee g)(x)\geqslant(f\vee g)(y)$，$(f\vee g)(x)=f(x)$，则有

$$|(f\vee g)(x)-(f\vee g)(y)|=f(x)-(f\vee g)(y)\leqslant f(x)-f(y).$$

其他三种情况是对称的，x 和 y 或者f和 g 互换，因此有

$$|(f\vee g)(x)-(f\vee g)(y)|\leqslant\max(|f(x)-f(y)|,\ |g(x)-g(y)|).$$

这就推出(a)的极大部分，极小部分是对称的．在(a)中用 $\|\cdot\|_\infty$取代 $\|\cdot\|_L$ 不等式是显然的，因此存在 i 和j，使得

$$\|f_1*\cdots*f_n\|_L\leqslant\|f_i\|_L,\qquad \|f_1*\cdots*f_n\|_\infty=\|f_j\|_\infty.$$

即证得(b). □

如果从$\mathbb{R}$映射到自身的函数f对于所有的x有导数$f'(x)$，那么可以证明f是一个利普希茨函数当且仅当f'是有界的(见习题1). 另一方面，$x\vee -x=|x|$，对所有的x，这里的每一个x和$-x$都是可微的(事实上，C^∞)，但是$|x|$在0处不是可微的，而是一个利普希茨函数.

对于有界利普希茨函数，我们有一个扩张定理.

11.2.3 命题 如果$A\subset S$和$f\in BL(A,d)$，那么f可以扩张为一个函数$h\in BL(S,d)$，满足在A上，$h=f$且$\|h\|_{BL}=\|f\|_{BL}$.

证明 根据 Kirszbraun-McShane 扩张定理(6.1.1)可知，存在S上的函数g，满足在A上，$g=f$且$\|g\|_L=\|f\|_L$. 令$h:=(g\wedge\|f\|_\infty)\vee(-\|f\|_\infty)$. 那么根据命题11.2.2(a)，在$A$上有$h=f$，且$\|h\|_L\leqslant\|g\|_L$，因此有$\|h\|_L\leqslant\|f\|_L$，$\|h\|_L=\|f\|_L$，所以$\|h\|_\infty\leqslant\|f\|_\infty$，从而$\|h\|_\infty=\|f\|_\infty$. □

回忆由于紧空间S上的连续函数是有界的，S上的所有连续实值函数构成的空间$C(S)$与$C_b(S)$是一致的. 这些函数可以用利普希茨来逼近.

391

11.2.4 定理 如果(S,d)是任意的紧度量空间，那么$BL(S,d)$在$C(S)$中对于$\|\cdot\|_\infty$是稠密的.

证明 根据命题11.2.1知，$BL(S,d)$是一个代数并且包含这个常量. 根据命题11.2.3知，它把A分成两个点集，f是其上的任意非常值函数，那么根据斯通–魏尔斯特拉斯定理(2.4.11)即可证得. □

例：如果$S=[0,1]$，并且f是S上的任意连续实值函数，令$f_n(x)=f(x)$，其中$x=j/n$，$j=0,1,\cdots,n$，并且令f_n在每个区间$[(j-1/n,j/n](j=1,\cdots,n)$上是线性的. 那么$f_n$是有界的和利普希茨的，并且当$n\to\infty$时，$\|f_n-f\|_\infty\to 0$.

11.2.5 推论 对于任意的紧度量空间(S,d)，$(C(S),\|\cdot\|_\infty)$是一个可分的巴拿赫空间.

证明 对每个$n=1,2,\cdots$，集合$\{f:\|f\|_{BL}\leqslant n\}$是一致有界和等度连续的，从而是完全有界的(实际上是紧的)，因此根据阿尔泽拉–阿斯科利定理(2.4.7)知，这个集合对于$\|\cdot\|_\infty$是可分的. 那么根据定理11.2.4知，这些集合的并在$C(S)$上是稠密的. □

例：如果S是有限维欧几里得空间$\mathbb{R}^k$上的一个紧集，那么根据魏尔斯特拉斯逼近定理(推论2.4.12)或它的推广——斯通–魏尔斯特拉斯定理(2.4.11)知，有理系数多项式形成了$C(S)$上的一个可数稠密集.

习题

1. 令f是$\mathbb{R}$映射到$\mathbb{R}$上的一个函数，其导数$f'(x)$对于所有的x都存在，证明：f是一个利普希茨函数，$\|f\|_L<\infty$当且仅当f'是有界的.
2. 证明：$\mathbb{R}$上的多项式是一个利普希茨函数当且仅当它的次数是0或1(它是常值或线性的)
3. 证明：$BL(S,d)$对于$\|\cdot\|_{BL}$是完备的，因此是一个巴拿赫空间.
4. 对于$S=\mathbb{R}$试举例证明：在命题11.2.2(b)中的常量2一般不能被任意更小的常量代替.
5. 对于$S=[0,1]$，其上的度量为通常的度量d，证明：$BL(S,d)$对于$\|\cdot\|_L$是不可分的，因此对于$\|\cdot\|_{BL}$是不可分的. [提示：考虑$f_t(x):=|x-t|$.]
6. 证明：命题11.2.2(a)中的不等式可以是严格的.
7. 给出推论11.2.5的等价证明，方法如下：令$\{x_n\}$是S上的一个稠密序列. $x\mapsto\{d(x,x_n)\}_{1\leqslant n<\infty}$将$S$映射到一个可数笛卡儿积上. 那么当$S$是具有积度量的区间的可数积时，此问题就可以根据命题2.4.4得到解决.

392

考虑由坐标点构成的多项式(每个多项式由有限多项组成)，其次再运用斯通–魏尔斯特拉斯定理(2.4.11).

8. (不动点定理). 令(S, d)是一个紧度量空间，F是从S映射到S的函数，满足对某个$t<1$，有$d(F(x), F(y))\leqslant td(x, y)$，$\forall x, y\in S$，令$x_0$是$S$上的任意点，由$x_{n+1}=F(x_n)(n=0, 1, \cdots)$递归来定义序列$\{x_n\}$.
   (a)证明：$\{x_n\}$是一个柯西序列.
   (b)令$x_n\mapsto y$. 证明：$F(y)=y$(y是F的一个不动点)是满足这个性质的S中的唯一点.
   注：如果$S=\mathbb{R}$且具有通常的度量，那么F上的假定是$\|F\|_L<1$.
9. 求解微分方程$dy/dx=G(x, y)$，其中G对于(x, y)是连续的，对y是利普希茨的，所以$\|G(x, \cdot)\|_L\leqslant K<\infty$，$\forall x\in[a, b]$. 对于$a\leqslant t\leqslant u\leqslant b$，令$S$是具有上确界距离的$C[t, u]$，由$F(f)(x)=\int_t^x G(y, f(y))dy$ $(t\leqslant x\leqslant u)$定义从$S$映射到自身的函数$F$.
   (a)t和u满足什么条件时，F满足习题8的条件?
   (b)证明：对任意的$c\in\mathbb{R}$，对$a\leqslant x\leqslant b$，存在唯一的连续函数f，使得$y=f(x)$满足微分方程，且有$f(a)=c$.
   [提示：用有限多个区间$[t, u]$覆盖$[a, b]$.]
10. 证明：当$x=0$时，微分方程$dy/dx=|y|^{\frac{1}{2}}$有不止一个解满足$y=0$.
11. 令f是从$\mathbb{R}$映射到$\mathbb{R}$的连续函数，使得对任意有限多个x，对某个$M<\infty$，$f'(x)$存在，且$|f'(x)|\leqslant M$，证明：$\|f\|_L\leqslant M$.
12. 证明：存在一个从$[0, 1]$映射到其自身的连续函数f，对勒贝格几乎所有的x有$|f'(x)|\leqslant 1$，但f不是一个利普希茨函数.[提示：在命题4.2.1中考虑康托尔函数.]

11.3 依 L 收敛的度量

对集合S的任意子集A，令A^c表示补集$S\setminus A$. (S, d)是一个度量空间，$\varepsilon>0$，令

393

$$A^\varepsilon := \{y\in S: \text{对某个}\ x\in A, d(x,y)<\varepsilon\}.$$

定义 对S上的任意两个法则P和Q，令

$$\rho(P,Q) := \inf\{\varepsilon>0: \text{对所有的博雷尔集}\ A, P(A)\leqslant Q(A^\varepsilon)+\varepsilon\}.$$

由于$A^\varepsilon=(\bar{A})^\varepsilon$，如果$A$是闭的，则可以得到一个等价的定义. 法则之间的距离ρ可以是S中A和A^ε的距离，并且通过加上ε使其在概率上是不同的.

例：对于两个点质量δ_x和δ_y(对任意x和集合A，$\delta_x(A):=1_A(x)$)，如果$d(x, y)\leqslant 1$，则有$\rho(\delta_x, \delta_y)=d(x, y)$(要证明$\rho(\delta_x, \delta_y)\leqslant d(x, y)$，分$x\in A$和$x\notin A$两种情况考虑；要证明边界对$d(x, y)\leqslant 1$可获得，取$A=\{x\}$). 如果$d(x, y)>1$，那么$\rho(\delta_x, \delta_y)=1$. 事实上，注意到，对任意法则$P$和$Q$，$\rho(P, Q)\leqslant 1$. 因此，$\rho(\delta_x, \delta_y)\equiv\min(d(x, y), 1)$.

11.3.1 定理 对于任意度量空间(S, d)，ρ是S中所有法则构成的集合上的一个度量.

证明 显然，对任意的P和Q，$\rho(P, Q)\geqslant 0$且$\rho(P, P)=0$.

如果$\rho(P, Q)>\varepsilon$，那么对某个博雷尔集A，有$P(A)>Q(A^\varepsilon)+\varepsilon$. 现在$A^{\varepsilon c\varepsilon}\subset A^c$，即如果对某个$y\in A^{\varepsilon c}$，$d(x, y)<\varepsilon$，则$x\notin A$，因为如果$x\in A$，则$y\in A^\varepsilon$. 因此，$Q(A^{\varepsilon c})>P(A^c)+\varepsilon\geqslant P(A^{\varepsilon c\varepsilon})+\varepsilon$. 从而$\rho(Q, P)=\rho(P, Q)$. 交换$P$和$Q$可以得到，$\rho(P, Q)=\rho(Q, P)$.

如果$\rho(P, Q)=0$，那么令A为任意闭集F，$\varepsilon=1/n$，其中$n=1, 2, \cdots$，注意到$F^{1/n}$的交是F，那么$P(F)\leqslant Q(F)$，同理，$Q(F)\leqslant P(F)$，所以$P(F)=Q(F)$. 从而根据闭的正则性可得$P=Q$(定理7.1.3).

如果M，P和Q是S上的法则，且$\rho(M, P)<x$和$\rho(P, Q)<y$，那么对任意的博雷尔集A，有

$$M(A)\leqslant P(A^x)+x\leqslant Q((A^x)^y)+y+x\leqslant Q(A^{x+y})+x+y,$$

所以 $\rho(M,\ Q)\leqslant x+y$. 令 $x\downarrow\rho(M,\ P)$ 和 $y\downarrow\rho(P,\ Q)$ 可得出，$\rho(M,\ Q)\leqslant\rho(M,\ P)+\rho(P,\ Q)$. □

度量 ρ 称为 Prohorov 度量(Prohorov metric)，或者有时称为 Lévy-Prohorov 度量．现在对 S 上的任意法则 P 和 Q，令 $\int f\mathrm{d}(P-Q):=\int f\mathrm{d}P-\int f\mathrm{d}Q$ 且

394

$$\beta(P,Q):=\sup\left\{\left|\int f\mathrm{d}(P-Q)\right|:\|f\|_{BL}\leqslant 1\right\}.$$

11.3.2 命题　对任意的度量空间 $(S,\ d)$，β 是 S 中所有法则构成的集合上的一个度量．

证明　根据定理 6.1.3，我们仅需要验证如果 $\beta(P,\ Q)=0$，那么 $P=Q$ 成立即可．对任意的闭集 F，如果 $g(x):=d(x,\ F)$，那么 $\|g\|_L\leqslant 1$(根据 2.5.3.)．因此对于函数 $f_m:=md(x,\ F)\wedge 1$，有 $\|f_m\|_{BL}\leqslant m+1$. 如果 U 是 F 的补集，由于 $f_m\uparrow 1_U$，我们可以得到 $P(U)=Q(U)$，$P(F)=Q(F)$ 和 $P=Q$. □

下面将证明 ρ 和 β 每个都度量化可分度量空间上的依 L 收敛．依 L 收敛的可度量化有下面的推论：假定法则 P_{nk} 和 P_n 满足 $P_n\to P_0$，且对每个 n，$P_{nk}\to P_n$. 那么存在一个子序列 $P_{nk(n)}$，使得 $P_{nk(n)}\to P_0$. 这不能直接得出，或者不容易从依 L 收敛的定义看出，部分是由于 $C_b(S)$ 一般不是可分的．

11.3.3 定理　对任意可分度量空间 $(S,\ d)$ 及 S 上的法则 P_n 和 P，下列各命题等价：

(a) $P_n\to P$.

(b) $\int f\mathrm{d}P_n\to\int f\mathrm{d}P$，$\forall f\in BL(S,\ d)$.

(c) $\beta(P_n,\ P)\to 0$.

(d) $\rho(P_n,\ P)\to 0$.

证明　(a)$\Rightarrow$(b) 显然成立．下面证明 (b)$\Rightarrow$(c)，令 T 是 S 的完备化空间，那么任意 $f\in BL(S,\ d)$ 可以唯一地扩张成 $BL(T,\ d)$ 上的一个函数，给出一个从 $BL(S,\ d)$ 映上到 $BL(T,\ d)$ 的 1－1 的线性映射，它保持范数 $\|\cdot\|_{BL}$. S 上的法则 P_n 和 P 同样可定义为 T 上的法则．从而在这一步可以假设 S 是完备的．那么根据 Ulam 定理(7.1.4)，对 $\forall\varepsilon>0$，取紧集 $K\subset S$，满足 $P(K)>1-\varepsilon$. 根据阿尔泽拉–阿斯科利定理(2.4.7)，将函数集合 $B:=\{f:\ \|f\|_{BL}\leqslant 1\}$ 限制在 K 上就形成了函数对于 $\|\cdot\|_\infty$ 的一个紧集，那么对某个有限的 k，存在 $f_1,\ \cdots,\ f_k\in B$，使得对任意的 $f\in B$，存在 $j\leqslant k$，满足 $\sup_{y\in K}|f(y)-f_j(y)|<\varepsilon$. 那么

$$\sup\{|f(x)-f_j(x)|:x\in K^\varepsilon\}<3\varepsilon,$$

由于如果 $y\in K$，且 $d(x,\ y)<\varepsilon$，则

395

$$\begin{aligned}|f(x)-f_j(x)|&\leqslant|f(x)-f(y)|+|f(y)-f_j(y)|+|f_j(y)-f_j(x)|\\&\leqslant\|f\|_L d(x,y)+\varepsilon+\|f_j\|_L d(x,y)<3\varepsilon.\end{aligned}$$

令 $g(x):=0\vee(1-d(x,\ K)/\varepsilon)$，则有 $g\in BL(S,\ d)$(运用 2.5.3 和命题 11.2.2)，且 $1_K\leqslant g\leqslant 1_{K^\varepsilon}$. 当 n 足够大时，$P_n(K^\varepsilon)\geqslant\int g\mathrm{d}P_n>1-2\varepsilon$.

于是对如上定义的 $\forall f\in B$ 和 f_j，对每个 $j=1,\ \cdots,\ k$，及足够大的 n，有

$$\begin{aligned}\left|\int f\mathrm{d}(P_n-P)\right|&\leqslant\int|f-f_j|\,\mathrm{d}(P_n+P)+\left|\int f_j\mathrm{d}(P_n-P)\right|\\&\leqslant 2(P_n+P)(S\setminus K^\varepsilon)+(3\varepsilon)\cdot 2+\left|\int f_j\mathrm{d}(P_n-P)\right|\end{aligned}$$

$$\leqslant 12\varepsilon + \left|\int f_j \mathrm{d}(P_n - P)\right| \leqslant 13\varepsilon$$

即可得(c).

接着证明(c)⇒(d). 给定一个博雷尔集 A 和 $\varepsilon>0$, 令

$$f(x) := 0 \vee (1 - \mathrm{d}(x,A)/\varepsilon).$$

那么 $\|f\|_{BL}\leqslant 1+\varepsilon^{-1}$. 对 S 上的任意法则 P 和 Q,

$$Q(A) \leqslant \int f\mathrm{d}Q \leqslant \int f\mathrm{d}P + (1+\varepsilon^{-1})\beta(P,Q) \leqslant P(A^\varepsilon) + (1+\varepsilon^{-1})\beta(P,Q).$$

那么 $\rho(P, Q)\leqslant\max(\varepsilon, (1+\varepsilon^{-1})\beta(P, Q))$, 因此, 如果 $\beta(P, Q)\leqslant\varepsilon^2$, 则有 $\rho(P, Q)\leqslant\varepsilon+\varepsilon^2$. 由于 $\rho(P, Q)\leqslant 1$, 可以得出 $\rho(P, Q)\leqslant 2\beta(P, Q)^{1/2}$, 即证得(d).

下面证明(d)⇒(a), 运用 portmanteau 定理(11.1.1), 令 A 为 P 的连续集, $\varepsilon>0$, 那么对 $0<\delta<\varepsilon$, 且 δ 充分小, 有 $P(A^\delta\setminus A)<\varepsilon$ 和 $P(A^{c\delta}\setminus A^c)<\varepsilon$. 那么当 n 充分大时, $P_n(A)\leqslant P(A^\delta)+\delta\leqslant P(A)+2\varepsilon$ 且 $P_n(A^c)\leqslant P(A^{c\delta})+\delta\leqslant P(A^c)+2\varepsilon$, 所以 $|(P_n-P)(A)|\leqslant 2\varepsilon$. 令 $\varepsilon\downarrow 0$, 即可得到 $P_n(A)\to P(A)$, 所以(a)得证. □

11.3.4 推论 令$(S, \mathcal{T})$是具有可数稠密子集的拓扑空间, 假设 S 上的法则 $P_n\to P$, 令$\mathcal{F}$是 S 上函数的任意一致有界等度连续族. 那么在$\mathcal{F}$上 P_n 一致收敛到 P, 换句话说,

$$\lim_{n\to\infty}\sup\left\{\left|\int f\mathrm{d}(P_n - P)\right| : f\in\mathcal{F}\right\} = 0.$$

注: "等度连续"意味着在每个点上等度连续. 在一个一般的拓扑空间(如这里的空间)上, "一致等度连续"甚至没有定义. 一个一致有界等度连续族的例子是度量空间上满足 $\|f\|_{BL}\leqslant 1$ 的所有函数 f 的集合$\mathcal{F}$. 另一方面, 任意有界连续函数的有限集是一致有界的和等度连续的, 无论这个函数对于给定的度量空间是不是利普希茨函数.

396

证明 令 $d(x, y):=\sup\{|f(x)-f(y)|: f\in\mathcal{F}\}$, $\forall x, y\in S$. 那么 d 是 S 上的伪度量, 即在除去那些 $x\neq y$ 但 $d(x, y)=0$ 的情况时是一个度量. 令 T 是 S 上满足$\{\langle x, y\rangle: d(x, y)=0\}$的点的等价类的集合. 那么存在一个从 S 映上到 T 的自然映射 G, 使得对 T 上的一个度量 e, 对 $\forall x, y$, 有 $d(x, y)=e(Gx, Gy)$. 对每个$f\in\mathcal{F}$, 对$\forall x, y$, 有 $|f(x)-f(y)|\leqslant d(x, y)$, 所以对于 d, $\|f\|_L\leqslant 1$. 同样, 如果 $d(x, y)=0$, 那么$f(x)=f(y)$. 因此$f(x)$仅依赖于 Gx: 对于 T 上的某个函数 h, 对 $\forall x$, $f(x)=h(G(x))$, 且对于 e, $\|h\|_L\leqslant 1$. 从而这些 h 构成的集合是等度连续和一致有界的. 因此可以假设 d 是一个度量.

由于$\mathcal{F}$等度连续, d 在 $S\times S$ 上是联合连续的, 这时(S, d)是可分的. 关于 d 定义利普希茨范数, 对 $\forall f\in\mathcal{F}$.

$$\|f\|_{BL} \leqslant 1 + \sup\{\|f\|_\infty : f\in\mathcal{F}\} < \infty$$

那么根据定理 11.3.3 中(a)⇒(c)的事实可以得出此结论. □

现在回忆对于随机变量由 $\alpha(X, Y):=\inf\{\varepsilon>0: P(d(X, Y)>\varepsilon)\leqslant\varepsilon\}$定义的樊畿度量, 它度量化依概率收敛(定理 9.2.2). 借助于 Prohorov 度量和樊畿度量, 依概率收敛可以推出依 L 收敛的事实(命题 9.3.5)能得到更明确的阐述.

11.3.5 定理 对任意的可分度量空间(S, d)和在概率空间上取值于 S 的随机变量 X 和 Y, $\rho(\mathcal{L}(X), \mathcal{L}(Y))\leqslant\alpha(X, Y)$.

证明 取 $\forall\varepsilon>\alpha(X, Y)$, 那么 $P(d(X, Y)\geqslant\varepsilon)<\varepsilon$. 任意的博雷尔集 A, 如果 $X\in A$ 和 $d(X,$

$Y)<\varepsilon$，那么 $Y\in A^{\varepsilon}$，所以

$$\mathcal{L}(X)(A) = P(X\in A) \leqslant P(Y\in A^{\varepsilon}) + \varepsilon = \mathcal{L}(Y)(A^{\varepsilon}) + \varepsilon.$$

因此 $\rho(\mathcal{L}(X),\ \mathcal{L}(Y))\leqslant\varepsilon$. 令 $\varepsilon\downarrow\alpha(X,\ Y)$ 即得证. □

在推论 11.6.4 中将要证明对完备可分度量空间上的任意两个法则 P 和 Q，存在随机变量 X 和 Y，满足 $\rho(P,\ Q)=\alpha(X,\ Y)$. 在这种意义下，定理 11.3.5 不能再改进. 397

习题

1. 证明：在 $\beta(P,\ Q)$ 的定义中，上确界仅限制于满足 $\|f\|_{BL}\leqslant 1$ 的那些 f，且有 $\sup f=\|f\|_{\infty}$ 和 $\inf f=-\|f\|_{\infty}$.
2. 对 $a<b$，令 $P_{a,b}$ 是区间 $[a,\ b]$ 上的一致分布，其关于勒贝格测度的密度为 $1_{[a,b]}/(b-a)$. 求
   (a)$\rho(P_{0,1},\ P_{0,2})$；(b)$\beta(P_{0,1},\ P_{0,2})$；(c)$\rho(P_{0,1},\ P_{1/2,3/2})$.
3. 给出定理 11.3.3 中(b)⇒(a)的另一种证明，运用 portmanteau 定理(11.1.1)的部分证明.
4. 假设 F 是 S 中的闭集，$x\geqslant 0$，$y>0$ 且 $P(F)>Q(F^{y})+x$. 证明：$\beta(P,\ Q)\geqslant 2xy/(2+y)$. [提示：证明存在一个函数 f，其在 F 上等于 1，在 F^{y} 外等于 -1，且有 $\|f\|_{L}\leqslant 2/y$. 注意到 $\int f\mathrm{d}(P-Q) = \int f+1\mathrm{d}(P-Q)$.]
5. 令 $g(x):=2x^{2}/(2+x)$，$x\geqslant 0$，
   (a)证明：对任意两个法则 P 和 Q，$g(\rho(P,\ Q))\leqslant\beta(P,\ Q)$. [提示：证明 g 是递增的，并且运用习题 4 的结论.]
   (b)推导
   $$\rho(P,Q) \leqslant \left(\frac{3}{2}\beta(P,Q)\right)^{1/2}.$$
   (c)证明：对满足 $0\leqslant t\leqslant 1$ 的任意 t，通过求 $\mathbb{R}$ 上的法则 P 和 Q，满足 $\rho(P,\ Q)=t$ 和 $\beta(P,\ Q)=g(t)$，证明：(a)部分的不等式是最佳的. [提示：令 P 是点质量 δ_0，令 Q 集中在 0 点和另外一个点.]
6. 考虑点式群体 $P:=\delta_p$ 和 $Q:=\delta_q$. 证明：当 $d(p,\ q)\to 0$ 时，$\rho(P,\ Q)/\beta(P,\ Q)\to 1$. 另外，对 $\mu:=(p+q)/2$，$\mu_n:=\mu+(P-Q)/n$，$d(p,\ q)=1/n$，证明：当 $n\to\infty$ 时，$\beta(\mu_n,\ \mu)/\rho(\mu_n,\ \mu)^{2}\to 1$.
7. 如同对实值函数那样，对复值函数定义 $\|f\|_{L}$ 和 $\|f\|_{\infty}$. 证明：对于概率测度的度量 $\beta'(P,\ Q)$ 定义为 β(其有复值函数)，事实上是等于 β 的. [提示：复值函数 f 可以通过乘以一个满足 $|z|=1$ 的常量 z 而使 $\int zf\mathrm{d}(P-Q)$ 成为实数.]
8. (莱维度量). 令 P 和 Q 是 $\mathbb{R}$ 上的两个法则，其分布函数分别为 F 和 G. 令 $\lambda(P,\ Q):=\inf\{\varepsilon>0$：对 $\forall x$，$F(x-\varepsilon)-\varepsilon\leqslant G(x)\leqslant F(x+\varepsilon)+\varepsilon\}$，
   (a)证明：λ 是可度量化 $\mathbb{R}$ 上依 L 收敛的度量. [提示：运用 Helly-Bray 定理(11.1.2).] 398
   (b)证明：$\lambda\leqslant\rho$，但存在法则 P_n 和 Q_n，使得 $\rho(P_n,\ Q_n)$ 不收敛于 0，而 $\lambda(P_n,\ Q_n)\to 0$.

11.4 经验测度收敛

对任意概率空间 $(S,\ \mathcal{B},\ \mu)$，存在概率空间 $(\Omega,\ P)$，在其上存在取值于 S 的独立随机变量 X_1，X_2，…，$\forall j$，$\mathcal{L}(X_j)=\mu$，因为可以把 Ω 看作是一 S 的笛卡儿积，把 X_j 作为坐标(8.2 节). 经验测度 μ_n 定义为

$$\mu_n(A)(\omega) := n^{-1}\sum_{j=1}^{n}\delta_{X_j(\omega)}(A), \quad A\in\mathcal{B},\omega\in\Omega.$$

如果 S 是任意拓扑空间，f 是 S 上有界连续实值函数，那么根据强大数定律(定理 8.3.5)，当 $n\to\infty$ 时，

$$\int f\mathrm{d}\mu_n = (f(X_1) + \cdots + f(X_n))/n \to \int f\mathrm{d}\mu \quad \text{a. s.}.$$

但是在概率为 0 的集合上该收敛不成立，它的收敛性可能依赖于f，并且有界连续函数的空间一般是不可分的，因此不能直接得到下面的事实．

11.4.1 定理(Varadarajan) 令(S, d)是可分度量空间，μ是S上的任意法则(博雷尔概率测度)．那么经验测度$\mu_n \to \mu$a. s.，即

$$P(\{\omega : \mu_n(\cdot)(\omega) \to \mu\}) = 1.$$

证明 证明时要用到如下事实(定理 2.8.2)：对于d拓扑来说，在S上存在一个完全有界的度量e. 例如，$\mathbb{R}$上的实直线对于它通常的度量显然不是完全有界的，但是$x \mapsto x$ arctan x是$\mathbb{R}$映射到有界开区间$[-\pi/2, \pi/2]$的同胚，它是完全有界的．因此可以假设(S, d)是完全有界的，根据定理 11.3.3，只需证明对所有的$f \in BL(S, d)$，$\int f(x)\mathrm{d}\mu_n(x)(\omega) \to \int f\mathrm{d}\mu$几乎必然成立．但是根据推论 11.2.5，$BL(S, d)$相对于上确界范数$\|\cdot\|_\infty$是可分的(对于范数$\|\cdot\|_{BL}$通常不是可分的)，由于$BL(S, d)$到$BL(C, d)$是等距的，这里的$C$是$S$的完备化空间并且是紧的．令$\{f_m\}$在$BL(S, d)$中对于$\|\cdot\|_\infty$是稠密的．那么根据强大数定律(定理 8.3.5)，对$\forall m$，$f = f_m$几乎必然成立，那么可以认为f_m接近于$f \in BL(S, d)$． □

399

在实值线$\mathbb{R}$上，令μ是其上的概率测度，其分布函数为

$$F(t) := \mu((-\infty, t]), \ -\infty < t < \infty.$$

那么经验测度μ_n有分布函数$F_n(t)(\omega) := \mu_n((-\infty, t])(\omega)$，相对于$F$，这里的$F_n$称为经验分布函数(empirical distribution function)．下面是一个经典的极限理论：

11.4.2 定理(Glivenko-Cantelli) 令μ是$\mathbb{R}$上的任意法则，其分布函数为F. 那么在$\mathbb{R}$上当$n \to \infty$时，$F_n(\cdot)(\omega)$几乎必然一致收敛到F.

证明 根据定理 11.4.1 和定理 11.1.2，对$\forall t$，几乎必然有$F_n(t) \to F(t)$，其中F在t处是连续的．在t值其余的至多可数集中，F可以有跳跃(定理 7.2.5)，把强大数定律(定理 8.3.5)运用到随机变量$1_{(-\infty, t]}(X_j)$，也可以有$F_n(t) \to F(t)$. 因此对所有的t，$F_n(t) \to F(t)$几乎必然成立．

为证明一致收敛，把一般情况归约为$[0, 1]$上勒贝格测度λ的情况上．令G是它的分布函数，对于所有的x，$G(x) = \max(0, \min(x, 1))$．

对于$\mathbb{R}$上的任意分布函数F，$0 < t < 1$，令$X_{F(t)} := \inf\{x : F(x) \geq t\}$. 回想对于$(0, 1)$上的勒贝格测度$\lambda$，根据命题 9.1.2，$X_F$是一个随机变量，其分布函数为$F$. 换句话说，$\lambda \circ X_F^{-1} = P$，其中$F$是$P$的分布函数．例如，如果$F$是$(2\delta_0 + \delta_2)/3$的分布函数，那么当$0 < t < 2/3$时，$X_F(t) = 0$，当$2/3 < t < 1$时，$X_F(t) = 2$. 下面的事实是有用的．

11.4.3 引理 如果G_n是G的经验分布函数，那么对于$\mathbb{R}$上的任意分布函数F，$G_n \circ F$是F的经验分布函数．

证明 令Y_j是 i. i. d. (λ)，因此，

$$G_n(F(x)) = n^{-1}\sum_{j=1}^{n} 1_{Y_j \leq F(x)}.$$

那么$X_j := X_F(Y_j)$是独立同分布的，根据命题 9.1.2 可知，该法则有分布函数F，同时也有$X_j \leq x$当且仅当$Y_j \leq F(x)$，所以该引理得证． □

继续证明定理 11.4.2，为了证明F_n几乎必然一致收敛到F，只需证明G_n几乎必然一致收敛到

G. 给定 $\varepsilon>0$，选择 m 足够大，使得 $1/m<\varepsilon/2$. 令

$$E:=\{k/m:k=0,1,\cdots,m\}.$$

那么在有限集 E 中，G_n 几乎必然收敛到 G，因此对于几乎所有的 ω，存在一个足够大的 n，使得 $\forall x\in E$，$|G_n-G|(x)(\omega)<\varepsilon/2$. 那么对于任意的 $x\in[0,1]$，取 $u, v\in E$，且 $u\leqslant x\leqslant v$，$u-v=1/m$. 即可得到 $G_n(x)\geqslant G_n(u)>u-\varepsilon/2>x-\varepsilon$，同理可得 $G_n(x)<x+\varepsilon$. 因此 $\forall x$，$|G_n-G|(x)<\varepsilon$. □

400

习题

1. 令 P 是区间$[0,2]$上的一致分布 $\lambda/2$，假设 $n=3$，$X_1=1.1$，$X_2=0.4$ 和 $X_3=1.7$. 给出 P 的分布函数 F 和经验分布函数 F_3. 并求$\sup_x|F_n-F|(x)$.
2. 证明：对 $\forall t$，存在满足 $H_n(t)\to H(t)$ 的分布函数 H_n，使得
   (a) H 不是一个分布函数，或者
   (b) H 是分布函数但 H_n 不一致收敛到 H. [提示：当 $t\leqslant 0$ 时，令 $H_n(t):=0$；当 $t>0$ 时，令 $H_n(t):=\min(t^n,1)$.]
3. 不运用命题 9.1.2 和引理 11.4.3 完成格利文科–坎泰利(Glivenko-Cantelli)定理(11.4.2)的证明. 运用逐点收敛且对 F 有跳跃的每个点 t，考虑开区间 $J_t:=(-\infty,t)$. 证明对所有这样的 t，$\mu_n(J_t)\to\mu(J_t)$ a. s.，然后运用这个结论完成定理的证明.
4. 对任意可分度量空间(S,d)和 S 上的法则 μ，证明：$\rho(\mu_n,\mu)\to 0$，$\beta(\mu_n,\mu)\to 0$a. s. 且当 $n\to\infty$ 时依概率收敛. [提示：主要的问题是证明 $\rho(\mu_n,\mu)$ 和 $\beta(\mu_n,\mu)$ 是可测随机变量，对于拓扑为 ρ 的 S 上的所有法则集$\mathcal{P}$(定理 11.3.3)，证明$(\mathcal{P},\rho)$是可分的，ρ 和 β 是$\mathcal{P}\times\mathcal{P}\mapsto\mathbb{R}$ 可测的，$\omega\mapsto P_n(\cdot)(\omega)$是从 Ω 到$\mathcal{P}$可测的.]
5. 令 P 是 S 上的法则.
   (a) 对于某个博雷尔集 F，尤其是有限或可数集，如果 $P(F)=1$，证明：对其他任意集 Q，$\rho(P,Q)$定义中的下确界可以在 $A\subset F$ 上取.
   (b) $\forall x\in S$ 和 $r>0$，令 $B(x,r):=\{y: d(x,y)<r\}$. 对于一个固定的 r，令 $f(x):=P(B(x,r))$. 证明：f 是上半连续的，即 $\forall t\in\mathbb{R}$，$\{y\in S: f(y)\geqslant t\}$是闭的.
   (c) 在习题 4 中，把(a)运用到 μ_n，(b)运用到 μ 得出 $\rho(\mu_n,\mu)$ 可测的另一个证明.
6. 如果 F 是有限集 $\{x_1,\cdots,x_n\}$，P 是法则，且 $P(F)=1$，证明：对任意集 Q，上确界 $\sup\{|\int fd(P-Q)|:\|f\|_L\leqslant 1\}$可以限制到形如 $f(x)=\min_{1\leqslant i\leqslant n}(c_i+d(x,x_i))$ 的函数上，其中 $c_1,\cdots,c_n$ 为有理数. [提示：利用命题 11.2.2.]

401

7. 假设 F 是$\mathbb{R}$ 上的连续分布函数，令 F_n 是其对应的经验分布函数. 对 $t\in\mathbb{R}$，求 $F_n(t)-F(t)$的方差的上确界. 令 $f_n=O_P(a_n)$，则 $\forall\varepsilon>0$，存在一个 $M<\infty$，使得 $\forall n$，$P(|f_n/a_n|>M)<\varepsilon$. 求 $\forall t$，使得 $F_n(t)-F(t)=O_P(n^c)$成立的最小的 c.

11.5 胎紧性和一致胎紧性

回忆胎紧概率测度的定义和法则的一致紧集(定理 9.3.3 前面). 本节首先将证明在任意适当的可测度量空间上，所有的法则都是胎紧的. 然后将证明一致胎紧性和法则的集合的紧性性质是紧密相关的.

定义 可分度量空间(S,d)是普遍可测的(universally measurable, u. m.)当且仅当对于 S 的完备化空间 T 上的每一个法则 P，存在博雷尔集 A 和 $B\in T$，满足 $A\subset S\subset B$ 和 $P(A)=P(B)$，使得 S 对于 P 的(测度论的)完备化是可测的(正如在 3.3 节所定义的).

例：如果 S 是 T 中的博雷尔集，那么显然 S 是普遍可测的. 因此大多数容易得到的空间 S 都是

普遍可测的．如果 S 在[0, 1]中不是勒贝格可测的(3.4 节)，那么它不是普遍可测的．这里存在普遍可测的集合但不是博雷尔集：例如，解析集$\{f(x): x\in S\}$，其中 f 是连续函数，S 是完备可分度量空间(见 13.2 节)．

11.5.1 定理 *可分度量空间(S, d)是普遍可测的，当且仅当 S 上的每个法则 P 都是胎紧的．*

证明 如果(S, d)是普遍可测的，令 P 是 S 的任意法则．那么对于完备化空间 T 中任意博雷尔集 A，在 T 上定义法则 $\bar{P}$ 为 $\bar{P}(A)=P(A\cap S)$，(这里的 $A\cap S$ 总是 S 中的博雷尔集，也可以看作是开集)，把 Ulam 定理(7.1.4)运用到 $\bar{P}$，并由普遍可测性知，在 S 上存在紧集 $K_n\subset S$，满足 $\bar{P}\left(\bigcup_{n\geq 1}K_n\right)=1$，那么 $P\left(\bigcup_{n\geq 1}K_n\right)=1$. 把集合放在更大的集合里不影响其紧性，因此，$P_n$ 是 S 中的紧集，P 是胎紧的．

[402] 相反，如果 S 上的每个法则都是胎紧的，令 Q 是 T 上的任意法则．定义 Q 的外测度：$Q^*(A):=\inf\{Q(B): B\supset A\}$，如果 $Q^*(S)=0$，那么 S 对于 Q 的完备化空间是可测的．如果$Q^*(S)>0$，对于 S 中的任意博雷尔集，令 $P(B):=Q^*(B)/Q^*(S)$. 那么根据定理3.3.6 知，P 是 S 上的法则．由于 P 是胎紧的，取紧集 $K_n\subset S$，且满足 $P(A)=1$，其中 $A=\bigcup_{n\geq 1}K_n$，那么$Q(A)=Q^*(S)$，因此，S 是普遍可测的．

□

11.5.2 推论 *(S, d)的普遍可测性质仅依赖于通过它拓扑的度量 d.*

事实上，普遍可测性质是由任何一个具有可测逆的 1－1(博雷尔)可测函数来保持的(见注释)，换句话说，它是由博雷尔同构保持的．另一方面，任意两个是其自身完备化空间的博雷尔子集的两个可分度量空间是博雷尔同构的，当且仅当他们有相同的基数，这将在定理 13.1.1 中证明．

在命题9.3.4 中已经证明了在$\mathbb{R}^k$ 中，每一个依 L 收敛的序列都是一致胎紧的．如果法则自身是胎紧的，则可以把它扩张到一般的度量空间上．

11.5.3 定理(Le Cam, 1957) *令(S, d)是一个度量空间，并且假设 $\forall n\geq 0$，P_n 是胎紧的，且 P_n 收敛到 P_0. 那么$\{P_n\}$是一致胎紧的．因此，如果 S 是可分的且是普遍可测的，那么每一个依 L 收敛的序列都是一致胎紧的．*

证明 每一个胎紧集 P_n 都集中在紧集 K_{nm}的可数并中．一个度量空间中的紧集是可分的(由于它是完全有界的)．因此所有 K_{nm}的并是可分的．令 T 是它的闭包，也是可分的．根据蒂策(Tietze)扩张定理(2.6.4)知，任意 T 中的有界连续实值函数都可以扩张为 S 上的有界连续函数．那么限制于 T 上的法则 P_n 仍收敛，因此根据定理 11.3.3 可知，它们对于 Prohorov 度量是收敛的，这些法则在 S 中也有同样的结论．

给定 $\varepsilon>0$，首先取一个紧集 K，满足 $P_0(K)>1-\varepsilon$. 对于每个 $n=1, 2, \cdots$，令

$$a(n):=\max(1/n, \inf\{\delta>0: P_n(K^{\delta})>1-\varepsilon\})$$

(对于11.3 节所定义的 K^{δ})．那么 $P_n(K^{2a(n)})>1-\varepsilon$. 由于对于所有的 Prohorov 度量，$P_n$ 都收敛到 P_0 (定理 11.3.3)，当 $n\to\infty$ 时，$a(n)\to 0$. 取 $K_n\subset K^{2a(n)}$，满足 $P_n(K_n)>1-\varepsilon$. 令 $L:=K\cup\bigcup_{n=1}^{\infty}K_n$. 序列

[403] $\{x_m\}\subset L$ 存在一个收敛子序列，如果 $x_m\in K$ 或者对于无穷多个 m，有一个特殊的 K_j. 否则，存在一个子序列 $x_{m(k)}\in K_{n(k)}$，$n(k)\to\infty$．那么对某些 $y_k\in K$，当 $k\to\infty$ 时，有 $d(x_{m(k)}, y_k)<2a(n(k))\to 0$. 存在一个子序列 $y_{(k(j))}\to y\in K$，所以 $x_{m(k(j))}\to y\in L$，因此$\{x_m\}$总有一个子序列在 L 中收敛，所以 L 是紧的．现在对 $\forall n$，$P_n(L)>1-\varepsilon$.

□

下面的定理把一致胎紧性和 11.3 节中讨论的法则的两个度量联系起来——Prohorov 度量 ρ 和对偶有界利普希茨度量 β.

11.5.4 定理　令 (S, d) 是完备的可分度量空间．令 A 为 S 上法则的集合，则下面各命题等价：

(Ⅰ) A 是一致胎紧的．

(Ⅱ) 对于 S 中的某个法则 P，A 中的每个序列 P_n 都有子序列 $P_{n(k)} \to P$.

(Ⅲ) 对于 S 上的所有法则的集合上的度量 β（或 ρ），A 有紧的闭包．

(Ⅳ) A 对于 β（或 ρ）是完全有界的．

注：条件(Ⅰ)和(Ⅱ)只依赖于 S 的拓扑，而不依赖特殊的度量．因此(Ⅰ)和(Ⅱ)在波兰空间（如 $S=(0, 1)$）上是等价的．另一方面，由于 S 不是完备的，(Ⅳ)不依赖于度量：例如，法则 $\delta_{1/n}$ 形成了 $(0, 1)$ 上对于 β 或 ρ 的完全有界集，正如在具有通常度量的 $[0, 1]$ 上，但是对于这两个度量，(Ⅰ)在 $(0, 1)$ 上不成立，然而在与 $(0, 1)$ 有同样的拓扑的 $\mathbb{R}$ 上，(Ⅰ)和(Ⅳ)是等价的．

证明　(Ⅱ)与(Ⅲ)的等价性可由可度量化定理(11.3.3)、紧性的等价性和度量空间中收敛子序列的存在性（定理 2.3.1）得到．显然(Ⅲ)⇒(Ⅳ).

(Ⅰ)⇒(Ⅱ)：假设 K_n 是紧的，$P(K_n) > 1-1/n$，$\forall P \in A$. 令 $\{P_m\}$ 是 A 中的序列．对于任意的 n，具有上确界范数的连续函数的空间 $C(K_n)$ 是可分的（推论 11.2.5）．对于 $S=\mathbb{R}^k$ 的证明（定理 9.3.3）可以运用到收敛子列 $P_{n(k)} \to P$. P 定义在使得任意 $f \in C_b(S, d)$ 可测的最小 σ-代数上，由于 (S, d) 是一个度量空间，因此，它是博雷尔 σ-代数．那么 P 是所需的法则．

最后将证明(Ⅳ)⇒(Ⅰ)．如果 A 对于 β 是完全有界的，那么 A 对于 ρ 也是完全有界的，由于 $\rho \leqslant 2\beta^{1/2}$，正如定理 11.3.3 的证明．因此假设 A 对于 ρ 是完全有界的．给定 $\forall \varepsilon > 0$，取有限集 $B \subset A$，使得对于 ρ，有 $A \subset B^{\varepsilon/2}$. 根据定理 11.5.1 或定理 7.1.4，每个 $P \in B$ 都是胎紧的，因此存在一个紧集 K_P，使得 $P(K_P) > 1-\varepsilon/2$. 令 K_B 是集合 K_P 的并，其中 $P \in B$，K_B 是紧集，且 $\forall P \in B$，$P(K_B) > 1-\varepsilon/2$. 取有限集 $F := F(\varepsilon) \subset S$，使得 $K_B \subset F^{\varepsilon/2}$. 所以 $\forall P \in B$，$P(F^{\varepsilon/2}) > 1-\varepsilon/2$. 那么根据 ρ 的定义可知，$\forall P \in A$，$P(F^{\varepsilon}) > 1-\varepsilon$. 然后 $\forall \delta > 0$，令 $\varepsilon(m) := \delta/2^m$，$m=1, 2, \cdots$，令 K 是 $F(\varepsilon(m))^{\varepsilon(m)}$ 闭包的交，那么 K 是紧的并且 $P(K) > 1-\delta$，$\forall P \in A$，所以 A 是一致胎紧的．　□

404

11.5.5 推论(Prohorov, 1956)　如果 (S, d) 是完备可分度量空间，那么 S 上所有法则的集合对于 ρ 和 β 是完备的．

证明　柯西序列（对度量 ρ 或 β）是完全有界的，因此根据定理 11.5.4 知，它有一个收敛子序列，所以也是收敛的．　□

习题

1. 令 X_j 是独立同分布变量，其分布为 $N(0, 1)$．令 H 是一个希尔伯特空间，规范正交基为 $\{e_j\}_{j\geqslant 1}$．令 $X := \sum_j X_j e_j / j$. 对每个 $\varepsilon > 0$，求一个紧集 $K \subset H$，使得 $P(X \in K) > 1-\varepsilon$.
2. 证明：在任意一个度量空间中，一个一致胎紧的法则序列有一个收敛子序列．
3. 证明：在度量空间 S 中，普遍可测子集的集族是一个 σ-代数.
4. 拓扑空间中集合 A 的一个点 x 称为孤立的，当且仅当 $\{x\}$ 在 A 的相对拓扑中是开的．一个紧集称为是完满的，当且仅当它不是孤立点．一个紧的完满度量空间在每一个点的邻域中都有不可数多个点（2.3 节习题 10）证明：对于任意的非空完满紧度量空间 K，在 K 的博雷尔集上存在一个法则 P，满足 $\forall x$，$P\{x\}=0$，和 $P(U)>0$，其中 U 是任意非空开集．［提示：寻找一个法则序列，其收敛子序列收敛到 P.］

5. 在[0, 1]上定义一个法则的特定有限集 F，使得对于[0, 1]上的每一个法则 P，存在 $Q\in F$，使得 $\rho(P, Q)<0.1$，这里的 ρ 是 Prohorov 度量．
6. 令 $C_0(\mathbb{R})$ 是 $\mathbb{R}$ 中所有连续实值函数 f 构成的集合，满足当 $|x|\to\infty$ 时，$f(x)\to 0$. 如果 μ_n 和 μ 是 $\mathbb{R}$ 上的有限测度，称 μ_n 淡收敛到 μ，如果 $\int f\mathrm{d}\mu_n\to\int f\mathrm{d}\mu, \forall f\in C_0(\mathbb{R})$. 证明：$\mathbb{R}$ 上的每一个法则序列有一个收敛子序列淡收敛到 $\mathbb{R}$ 中的某个测度 μ，其中 $0\leqslant\mu(\mathbb{R})\leqslant 1$.［提示：见 2.8 节的习题 1(a).］

405

7. 令 H 是无穷维可分希尔伯特空间，其规范正交基为 $\{e_n\}_{n\geqslant 1}$. 令 $e(n):=e_n$ 和 $P_n:=\delta_{e(n)}$. 对于任意的子列 $P_{n(k)}$，求 H 中的一个有界连续函数 f，使得 $\int f\mathrm{d}P_{n(k)}$ 不收敛．
8. 令 (S, d) 是一个度量空间，令 A 是 S 上的一个一致胎紧的法则集．证明：A 对于 ρ 和 β 有紧的闭包．［提示：见习题 2.］
9. 称拓扑空间中的一个集合 A 普遍测度为 0，当且仅当对 A 的博雷尔集上的每个测度 μ，其中 μ 是非原子的（即对每个点 x，$\mu\{x\}=0$），$\mu^*(A)=0$. 在连续统假设条件下，证明：存在一个不可数集 A，其在[0, 1]上普遍测度为 0.［提示：用等价性定理 1.4.1 证明在[0, 1]的博雷尔 σ-代数上存在 c 个非原子法则．在超限递归(1.3.2)中，令 μ_2 为非原子法则．交替的取新的点 $x_\alpha\subset A$ 和 $B_\alpha\subset A^c$，其中 B_α 是第一范畴集，其与使 $\mu_\alpha\left(\bigcup_{\beta\leqslant\alpha}B_\beta\right)=1$ 成立的可数集 $\{x_\beta\}_{\beta\leqslant\alpha}$ 不相交．］

11.6 斯特拉森定理：具有邻近法则的邻近变量

斯特拉森(Strassen)定理指出，如果两个法则在 Prohorov 度量 ρ 下是接近的，那么存在具有这些法则的随机变量 X 和 Y，其中法则是依概率等邻近的（对于樊畿度量）．换句话说，(X, Y) 的联合分布集中在对角线 $X=Y$ 上（所以 X 和 Y 不是独立的）．其证明是基于下面一个有穷组合的事实，称为配对定理．

给定两个集合 A 和 B，关系是子集 $K\subset X\times Y$. 那么 xKy 就意味着 $\langle x, y\rangle\in K$. 对任意的 $A\subset X$，令 $A^K:=\{y\in Y: xKy, x\in A\}$. A 映射到 B 的 K-配对 f 是一个从 A 映射到 B 的 $1-1$ 函数，满足 $\forall x$，$xKf(x)$. 对任意有限集 A，$\mathrm{card}(A)$（A 的势）就是指 A 中元素的数目．

11.6.1 配对定理(D. König-P. Hall) *令 X 和 Y 是有限集，满足关系 $K\subset X\times Y$，并且对任意的集合 $A\subset X$，有 $\mathrm{card}(A^K)\geqslant\mathrm{card}(A)$. 那么存在一个从 X 映射到 Y 的 K-配对 f，因此如果 $\mathrm{card}(Y)=\mathrm{card}(X)$，则映上到 Y.*

注：*给出的 K-配对存在的充分条件也是必要条件，由于对于每个 A，f 必须是从 A 映射到 A^K 的 $1-1$ 函数．该定理也称为“婚配引理”：如果 X 是女性的集合，Y 是男性的集合，xKy 就意味着 x 和 y 是相容的，那么定理给出了 x 能够与 y 相容的精确条件．*

406

证明 令 $\mathrm{card}(X)=m$. 这个证明是通过对 m 作归纳来证得的．对于 $m=1$，结论是显然的．假设结论对于 $1, \cdots, m-1$ 成立．选择某个 $x\in X$，那么对于某个 $y\in Y$，有 xKy. 如果存在一个从 $X\setminus\{x\}$ 映射到 $Y\setminus\{y\}$ 的 K-配对 f，则令 f 为 $f(x)=y$. 否则，由归纳假设知，存在一个集合 $A\subset X\setminus\{x\}$，满足 $\mathrm{card}(A^K\setminus\{y\})<\mathrm{card}(A)$，因此 $\mathrm{card}(A^K)=\mathrm{card}(A)>0$. 并且存在一个 A 映上到 A^K 的 K-配对 f. 如果假设条件对于 $X\setminus A$ 和 $Y\setminus A^K$ 成立（限定关系 K 为这些集合的积），我们可以定义所有 X 上的配对．否则，存在集合 $D\subset X\setminus A$，使得 $\mathrm{card}(D^K\setminus A^K)<\mathrm{card}(D)$. 但是

$$\mathrm{card}((A\cup D)^k)=\mathrm{card}(A^K)+\mathrm{card}(D^K\setminus A^K)<\mathrm{card}(A)+\mathrm{card}(D)=\mathrm{card}(A\cup D),$$

矛盾． □

令 (S, d) 是可分度量空间，$\mathcal{P}(S)$ 表示 S 中所有法则的集合．给定 $A\subset S$ 和 $\delta>0$，回忆在

Prohorov 度量空间的定义(11.3 节)中定义的

$$A^{\delta} = \{x: 对某个 y \in A, d(x,y) < \delta\} = \{x: d(x,A) < \delta\},$$

并且 $d(\cdot, A)$ 是(2.5.3)中所定义的．那么 A^{δ} 是开的．定义闭集 $A^{\delta]}$ 为 $A^{\delta]} := \{x: d(x, A) \leqslant \delta\}$.

在笛卡儿积 $A \times B$ 中，坐标投影 $p_1(\langle x, y\rangle) := x$ 和 $p_2(\langle x, y\rangle) := y$. 对于 $A \times B$ 上的一个法则 μ，它的边缘(marginal)是 A 上的法则 $\mu \circ p_1^{-1}$ 和 B 上的法则 $\mu \circ p_2^{-1}$. 例如，如果 μ 是一个积测度 $P \times Q$，其中 P 和 Q 是法则，那么 P 和 Q 是边缘．另一方面，如果 P 是 A 上的法则，μ 是 $A \times A$ 上的对角法则 $P \circ D^{-1}$，其中 $D(x) := \langle x, x\rangle$，那么 μ 的两个边缘都等于 P.

定理 11.3.5 表明两个法则之间的 Prohorov 距离的上界为具有这两个法则的任意两个随机变量之间的樊畿距离．另一方面，具有同一法则的两个随机变量可能是独立的，因此对于樊畿度量不是接近的．另外一个事实是定理 11.3.5 的逆命题，它将要证明(对于 $\alpha = \beta$)对于任意两个法则，存在具有这些法则的随机变量，如同法则在 Prohorov 度量那样在樊畿度量中是接近的(或者对于不完备空间几乎是这样的).

11.6.2 定理　对于任意可分度量空间(S, d)，S 中的任意两个法则 P 和 Q，$\alpha \geqslant 0$ 和 $\beta \geqslant 0$，下列两个命题等价：

(Ⅰ)对任意闭集 $F \subset S$，$P(F) \leqslant Q(F^{\alpha]}) + \beta$.

(Ⅱ)对任意的 $a > \alpha$，存在一个法则 $\mu \in \mathcal{P}(S \times S)$，其边缘是 P 和 Q，使得 $\mu\{\langle x, y\rangle: d(x, y) > a\} \leqslant \beta$.

如果 P 和 Q 是胎紧的，那么在(Ⅱ)中对于 $a = \alpha$，μ 是可选取的．

407

注：如果 $\alpha = \beta$，那么(Ⅰ)表明 $\rho(P, Q) \leqslant \alpha$，其中 ρ 是 Prohorov 度量．S 中的所有法则是胎紧的，如果它是完备的(Ulam 定理 7.1.4)或仅仅是普遍可测的(定理 11.5.1)．本节习题 5 就是证明对于一个一般(不可测)的集合 S，不能取 $a = \alpha$. 定理 11.6.2 将要由下面更详细的事实来证明．

11.6.3 定理　假设定理 11.6.2 中(Ⅰ)，对于任意的 $a > \alpha$ 和 $b > \beta$，存在 $S \times S$ 上的非负博雷尔测度 η 和 γ，使得

(1)$\eta + \gamma$ 是 $S \times S$ 中具有 S 上的边缘 P 和 Q 的法则．

(2)η 集中在$\langle x, y\rangle$集合中，满足 $d(x, y) \leqslant a$.

(3)γ 有小于等于 b 的全质量．

(4)$\eta + \gamma$ 都是积测度的有限和．

证明　取 $\varepsilon > 0$，使得 $\varepsilon < \min(a - \alpha, b - \beta)/5$. 首先证明一些简单的情况．

情况 A：存在整数 n 和集合 $M \subset S$，$N \subset S$，且 $\mathrm{card}(M) = \mathrm{card}(N) = n$，使得 $\forall x \in M$，$y \in N$，$P\{x\} = Q\{y\} = 1/n < \varepsilon$. 取整数 k，满足 $n\beta < k < n(\beta + \varepsilon)$. 取有 k 个元素的集合 U 和 V，其与前面提到的所有集合不相交．令 $X := M \cup U$ 和 $Y := N \cup Y$. 对 $x \in X$ 和 $y \in Y$，定义 xKy 当且仅当：

(a)$x \in U$，或(b)$y \in V$，或(c)$d(x, y) \leqslant \alpha$，$x \in M$ 和 $y \in N$. 给定 $A \subset X$，满足 $\mathrm{card}(A) = r$，为证明 $\mathrm{card}(A^K) \geqslant r$ 可以假设 $A \subset M$，由于否则 $A^K = Y$. 那么

$$r/n = P(A) \leqslant Q(A^{\alpha]}) + \beta \leqslant \beta + \mathrm{card}(A^K \cap N)/n,$$

所以

$$r \leqslant n\beta + \mathrm{card}(A^K \cap N) < k + \mathrm{card}(A^K \cap N) = \mathrm{card}(A^K).$$

因此运用配对定理(11.6.1)，考虑 X 映上到 Y 的一个 K-配对 f. 那么对 M 中至少 $n - k$ 个元素 x，$f(x) \in N$，形成一个集合 T，那么 $d(x, f(x)) \leqslant \alpha$. 对于这样的 x，令 $g(x) = f(x)$，把 g 扩张成

从 M 映射到 N 的 1－1 函数．令

$$\mu := \sum_{x\in M}\delta_{\langle x,g(x)\rangle}/n,\eta := \sum_{x\in T}\delta_{\langle x,g(x)\rangle}/n, \qquad \gamma := \mu - \eta.$$

那么 $\mu \in \mathcal{P}(S\times S)$，且有边缘 P 和 Q，η 和 γ 有性质(1)～(4)．

情况 B：P 和 Q 每个都集中在有限多个点上并且给出一个有理概率．那么对于满足 $1/n<\varepsilon$ 的某个正整数 n，P 和 Q 在 $\{j/n: j=0, 1, \cdots, n\}$ 中取值，令 J 是有 n 个元素的集合(与 S 不相交)，比如 $J=\{1, 2, \cdots, n\}$. 对 $i, j\in J$，令 $f(i, j)=\begin{cases}\varepsilon, & 若\ i\neq j\\ 0, & 若\ i=j\end{cases}$. 在 $S\times J$ 上，令

$$e(\langle x,i\rangle,\langle y,j\rangle) := d(x,y)+f(i,j)\leqslant d(x,y)+\varepsilon.$$

那么 e 是一个度量．定义一个法则 $P_1\in\mathcal{P}(S\times J)$，使得对 P 的每一个原子 x，有 $P\{x\}=j/n$，$P_1(\langle x, i\rangle)=\begin{cases}1/n, & 若\ i=1, \cdots, j\\ 0, & 若\ i>j\end{cases}$. 那么 P_1 在 S 上有边缘 P. 令 Q_1 是 $S\times J$ 上类似的法则，其在 S 上有边缘 Q.

用 $\alpha+\varepsilon$ 代替 α，情况 A 的假设对于 P_1 和 Q_1 成立．所以根据情况 A，在 $(S\times J)\times(S\times J)$ 上有法则 $\mu_1=\eta_1+\gamma_1$，其在 $S\times J$ 上有边缘 P_1 和 Q_1，使得 η_1 和 γ_1 对于 μ_1 有性质(1)～(4). 令 μ，η 和 γ 分别是 μ_1，η_1 和 γ_1 在 $S\times S$ 上的边缘．那么由于 $\alpha+\varepsilon<a$，η 和 γ 有性质(1)～(4).

情况 C：这是一个一般的情况．给定 $\varepsilon>0$，令 A 是满足 $d(x, y)\geqslant\varepsilon(\forall x\neq y\in A)$ 成立的 S 的最大子集，那么 A 是可数的，设 $A=\{x_j\}_{j\geqslant1}$，可能是有限的．令

$$B_j := \{x\in S: d(x,x_j)<\varepsilon\leqslant d(x,x_i), i<j\}.$$

那么 B_j 是与 S 的并不相交的博雷尔集．定义法则 $P'\{x_j\}=p_j:=P(B_j)$，$j=1, 2, \cdots$. 同理由 Q 定义 Q'. 那么对任意博雷尔集 A，有 $P'(A)\leqslant P(A^\varepsilon)$. 首先假设 $A\neq S$，取 $x_0\in S\setminus A$. 对某个整数 n，令 $P''\{x_j\}=[np_j]/n$，$\forall j\geqslant1$，其中 $[x]$ 是指小于等于 x 的最大整数．定义 $P''\{x_0\}:=1-\sum_{j\geqslant1}P''\{x_j\}$，同理由 Q 定义 Q''. 取足够大的 n，使得 $P''\{x_0\}<\varepsilon$ 和 $Q''\{x_0\}<\varepsilon$. 那么对任意闭集 F，由于 $(F^\varepsilon)^\alpha\subset F^{\varepsilon+\alpha}$，并且闭集当 $\delta\uparrow\varepsilon$ 时，$F^{\delta]}\uparrow F^\varepsilon$，

$$\begin{aligned}P''(F) &\leqslant P'(F)+\varepsilon\leqslant P(F^\varepsilon)+\varepsilon\leqslant Q(F^{\varepsilon+\alpha})+\beta+\varepsilon\\ &\leqslant Q'(F^{2\varepsilon+\alpha})+\beta+\varepsilon\leqslant Q''(F^{2\varepsilon+\alpha})+\beta+2\varepsilon,\end{aligned}$$

其中第二个至最后一个不等式由 $Q(C)\leqslant Q'(C^\varepsilon)$($C$ 为任意博雷尔集)得到，由于如果 C 与 B_j 相交，C^ε 包含 x_j，所以对任意博雷尔集 C，$Q(C)\leqslant Q'(C)$. 现在将情况 B 运用到 P''和 Q''，$\mu''=\eta''+\gamma''$在 $S\times S$ 上有边缘 P''和 Q''，$\eta''(d>3\varepsilon+\alpha)=0$ 和 $\gamma''(S\times S)\leqslant\beta+3\varepsilon$. 这些测度集中在点 $\langle x_i, y_j\rangle$ 上，其质量都是 $1/n$ 的倍数．如果通过包含它的所有在 $\langle x_0, x_j\rangle(\forall j)$ 或 $\langle x_i, x_0\rangle(\forall i)$ 上的质量来改变 γ'' 的定义，那么除了 $\gamma''(S\times S)\leqslant\beta+5\varepsilon$ 外，上面所有的结论都成立．在特殊的情况($S=A$)下，对于 $A\cup\{x_0\}(x_0\notin A)$ 我们证明上述结论，然后在使 A 不动的映射下取像测度并取 x_0 到 x_1.

对任意博雷尔集 C，令 $P_i(C):=P(C\mid B_i):=P(C\cap\mid B_i)/P(B_i)$，或者如果 $P(B_i)=0$，令 $P_i:=0$，$i=1, 2, \cdots$. 同理由 Q 定义 Q_j，$j=1, 2, \cdots$. 令

$$\eta''=\sum_{i,j}\frac{k(i,j)}{n}\delta_{\langle x_i,x_j\rangle}, \qquad \eta := \sum_{i,j}\frac{k(i,j)}{n}(P_i\times Q_j)$$

其中 $k(i, j)$ 是非负整数，对 $i, j\geqslant1$ 求和，并且 $d(x_i, x_j)\leqslant3\varepsilon+\alpha$. 令 η 有边缘 u 和 v. 那么可以验证 $u\leqslant P$，同理 $v\leqslant Q$. 令 $t:=(P-u)(S)=(Q-v)(S)\leqslant\beta+5\varepsilon$. 如果 $t=0$，令 $\gamma:=0$，否则，令 $\gamma:=(P-u)\times$

$(Q-v)/t$. 那么显然 $\eta+\gamma$ 是有边缘 P 和 Q 的法则，且(1)～(4)成立．即证得定理 11.6.3. □

定理 11.6.2 的证明 要证明(Ⅱ)⇒(Ⅰ)，对于满足(Ⅱ)的 μ 和任意的闭集 F,

$$\begin{aligned}P(F) &= \mu(F\times S) \leqslant \beta+\mu\{\langle x,y\rangle : x\in F, d(x,y)\leqslant a\}\\ &\leqslant \beta+\mu\{\langle x,y\rangle : y\in F^{a}\} = \beta+Q(F^{a]}).\end{aligned}$$

令 $a=a(n):=\alpha+1/n$, $n=1,2,\cdots$. 那么 $\bigcap_n F^{a(n)}=F^{\alpha]}$，即证得(Ⅰ).

反之，要证(Ⅰ)⇒(Ⅱ)，首先假设 P 和 Q 是胎紧的．根据定理 11.6.3，对 $n=1,2,\cdots$存在 $S\times S$ 上的法则 μ_n，对于所有的 n，其有边缘 P 和 Q，使得

$$\mu_n(d>\alpha+1/n\} < \beta+1/n, n=1,2,\cdots.$$

如果 K 和 L 是紧的，且 $P(K)>1-\varepsilon$, $Q(L)>1-\varepsilon$，那么对 $\forall n$, $\mu_n((S\times S)\setminus(K\times L))\leqslant 2\varepsilon$，所以$\{\mu_n\}$是一致胎紧的，并且根据定理 11.5.4 知，有一收敛子序列收敛到 μ, μ 有边缘 P 和 Q. 那么根据 portmanteau 定理(11.1.1)，对每个 n, $\mu(d>\alpha+1/n)\leqslant\beta$，所以 $\mu(d>\alpha)\leqslant\beta$，即在这种情况下证得定理．

现在证明一般情况，令 $\bar{S}$ 是 S 的完备化空间．对于 $\bar{S}$ 中的任意博雷尔集 C，令 $\bar{P}(C):=P(C\cap S)$. 那么 $\bar{P}$ 是 $\bar{S}$ 上的一个法则．同理在 $\bar{S}$ 上定义 $\bar{Q}$. 根据 Ulam 定理(7.1.4)，$\bar{P}$ 和 $\bar{Q}$ 是胎紧的．对于 $\bar{S}$ 上的任意闭集 F，有

$$\bar{P}(F) = P(F\cap S) \leqslant Q((F\cap S)^{\alpha]}+\beta) \leqslant \bar{Q}(F^{\alpha]})+\beta.$$

因此根据胎紧的情况，在 $\bar{S}\times\bar{S}$ 上存在一个法则 ν，其有边缘 $\bar{P}$ 和 $\bar{Q}$，满足 $\nu(d>\alpha)\leqslant\beta$. 如同定理 11.6.3 证明中的情况 C，给定 $0<\varepsilon<(a-\alpha)/2$，令 $\bar{S}$ 是不相交博雷尔集 $\bar{B}_j$ 的并，其中 $\bar{B}_j$ 满足对每个 j, $\forall x, y\in\bar{B}_j$, $d(x,y)\leqslant\varepsilon$. 令 $c_{jk}:=\nu(\bar{B}_j\times\bar{B}_k)$, $p_j:=\bar{P}(\bar{B}_j)=P(B_j)$ 和 $q_k:=\bar{Q}(\bar{B}_k)=Q(B_k)$, $j,k=1,2,\cdots$. 那么 $p_j\equiv\sum_k c_{jk}, q_k\equiv\sum_j c_{jk}$. P_j 和 Q_k 如以前所定义．令 $\mu:=\sum_{j,k}c_{jk}(P_j\times Q_k)$，对任意博雷尔集 $D\subset\bar{S}\times\bar{S}$，令 $\bar{\mu}(D):=\mu(D\cap(S\times S))$. 那么 μ 有边缘 P 和 Q，对 $\forall j, k$, $\bar{\mu}(\bar{B}_j\times\bar{B}_k)=c_{jk}$,

$$\begin{aligned}\mu(d>\alpha+2\varepsilon) &= \bar{\mu}(d>\alpha+2\varepsilon)\\ &\leqslant \sum_{j,k}\{c_{jk}: d(y,z)>\alpha,\quad \forall y\in\bar{B}_j,\quad z\in\bar{B}_k\}\\ &\leqslant \nu(d>\alpha)\leqslant\beta,\end{aligned}$$

即证得定理 11.6.2. □

现在对于随机变量，樊畿度量 α 度量化依概率收敛(定理 9.2.2)，它与法则的 Prohorov 度量有如下关系．

11.6.4 推论 对任意可分度量空间(S,d)，S 上的法则 P 和 Q，$\varepsilon>0$ 或 $\varepsilon=0$，如果 P 和 Q 是胎紧的，存在一个概率空间(Ω,μ)和 Ω 上的随机变量 X, Y，满足

$$\alpha(X,Y)\leqslant\rho(P,Q)+\varepsilon,\qquad \mathcal{L}(X)=P, \text{and } \mathcal{L}(Y)=Q.$$

所以

$$\rho(P,Q)=\inf\{\alpha(X,Y):\mathcal{L}(X)=P,\mathcal{L}(Y)=Q\}.$$

证明 对于第一部分，运用定理 11.6.2，其中 $\alpha=\beta=\rho(P,Q)$，一般地 $\alpha<a<\alpha+\varepsilon$，或者如果 P 和 Q 是胎紧的，$a=\alpha$. 由于总是有 $\rho(\mathcal{L}(X),\mathcal{L}(Y))\leqslant\alpha(X,Y)$(定理 11.3.5)，令 $a\downarrow\alpha(X,Y)$即可得出结论． □

现在回忆对偶有界利普希茨度量 β 和 Prohorov 度量 ρ 度量化法则上的同一拓扑(定理 11.3.3).

410

更具体地，在定理 11.3.3 的证明中，(c)⇒(d)证得 $\rho\leqslant 2\beta^{1/2}$. 下面是在其他方向上的不等式.

11.6.5 推论 对任意可分度量空间 S 以及 S 上的法则 P 和 Q，有 $\beta(P,Q)\leqslant 2\rho(P,Q)$.

证明 给定 $\varepsilon>0$，按推论 11.6.4 取随机变量 X 和 Y，令

$$A:=\{\langle x,y\rangle\in S\times S: d(x,y)\leqslant\rho(P,Q)+\varepsilon\}.$$

411 取任意的 $f\in BL(S,d)$，那么

$$\left|\int f\mathrm{d}(P-Q)\right| = |Ef(X)-Ef(Y)|\leqslant E|f(X)-f(Y)|$$
$$\leqslant [\|f\|_L E1_A(X,Y)+2\|f\|_\infty](\rho(P,Q)+\varepsilon)$$
$$\leqslant 2\|f\|_{BL}(\rho(P,Q)+\varepsilon).$$

令 $\varepsilon\downarrow 0$ 即证得. □

接着 ρ 和 β 在 S 的所有法则的集合上定义了同样的一致结构，一致性将在下节中证明.

习题

1. 令 $n_{i,j}=0$ 或 1，$i=1,\cdots,k$ 和 $j=1,\cdots,m$，所以 $\{n_{i,j}\}$ 形成了一个元素为 0 或 1 的 $k\times m$ 长方矩阵，称对 $\langle A,B\rangle$(其中 $A\subset\{1,\cdots,k\}$，$B\subset\{1,\cdots,m\}$)是一个覆盖(cover)，如果当 $n_{ij}=1$ 时，或者 $i\in A$ 或者 $j\in B$. 对于所有覆盖 $\langle A,B\rangle$，令 c 是基数 card(A)+card(B)的最小和. 令 d 是集合 $D\subset\{1,\cdots,k\}\times\{1,\cdots,m\}$ 的最大基数且使得 $\forall\langle i,j\rangle\in D$ 和 $\langle i,j\rangle\neq\langle r,s\rangle\in D$，$i\neq r$ 且 $j\neq s$，$n_{ij}=1$. 证明 $c=d$. [提示：运用定理 11.6.1.]
2. 证明：推论 11.6.5 在 $\mathbb{R}$ 中其结论中的不等式是严格的，在此情形中，对于 $0\leqslant t\leqslant 1$，$\sup\{\beta(P,Q):\rho(P,Q)=t\}=2t$. [提示：令 P 和 Q 每个都有大小为 t 的原子，并且二者在所有点处有较大距离，否则 P 和 Q 相同.(另一种思路可以参见 11.3 节习题 5.)]
3. (a)证明：对于任意紧集 K 和 $\delta>0$，如果 $y\in K^{\delta]}$，那么 $d(x,y)\leqslant\delta$，$x\in K$.
   (b)证明：在无限维希尔伯特空间中，存在一个闭集 F，其有离散相对拓扑，$y\in F^{1]}$ 且对 $\forall x\in F$，$d(x,y)>1$. [提示：对于规范正交的 e_n 和某个 a_n，令 $F=\{a_ne_n\}_{n\geqslant 1}$.]
4. 在 $\mathbb{R}$ 中，给定一个概率分布函数 F，回忆关于(0，1)上的勒贝格测度 λ，随机变量 X_F 定义为 $X_F(t):=\inf\{x: F(x)\geqslant t\}$，其中 $0<t<1$(命题 9.1.2). 试举例：两个法则 P 和 Q，分别具有分布函数 F 和 G，使得 $\alpha(X_F,X_G)>\rho(P,G)$，因此，在推论 11.6.4 中不能取 $X=X_F$ 和 $Y=X_G$(ε 足够小).
5. 令 E 是[0，1]中的一个集合，其有勒贝格内测度为 0，外测度为 1(定理 3.4.4)，即 $\lambda_*(E)=0$ 和 $\lambda^*(E)=1$. 412 令 $T:=\{x+1: x\in[0,1]\setminus E\}$ 和 $S:=E\cup T$. 定义 S 中的法则 P 和 Q：$P(A):=\lambda^*(A\cap E)$，$Q(A):=\lambda^*(A\cap T)$，根据定理 3.3.6，S 的子集 A 是一个博雷尔集(一般情况下不是 $\mathbb{R}$ 中的博雷尔集). 证明：对于 $\alpha=1$ 和 $\beta=0$，定理 11.6.2 中的条件(Ⅰ)成立，而对于 $\alpha=1$ 和 $\beta=0$ 条件(Ⅱ)不成立，证明没有胎紧性可能取不到 $a=\alpha$.
6. 在定理 11.6.3 中，假设 $\gamma(S\times S)>0$. 令 γ' 是积测度 $\gamma'=(\gamma\circ p_1^{-1})\times(\gamma\circ p_2^{-1})/\gamma(S\times S)$. 证明或反正：当 γ 用 γ' 代替时，性质(1)～(4)成立.
7. 回忆 10.2 节中所定义的条件分布 $P_{\cdot|\cdot}(\cdot,\cdot)$，令 $(X,\mathcal{A})$ 和 $(Y,\mathcal{B})$ 是两个可测空间. 令 P 是在 $X\times Y$ 的积 σ-代数上的一个概率测度. 令 $h(x,y)\equiv y$. 令 $\mathcal{C}$ 是 $X\times Y$ 子集的最小 σ-代数，其中 x 是可测的，因此 $\mathcal{C}=\{A\times Y: A\in\mathcal{A}\}$. 假设 P 是积测度的有限和. 证明：存在一个在 $\mathcal{B}\times(X\times Y)$ 中的条件分布函数 $P_{h|\mathcal{C}}(\cdot|\cdot)$，其中 $P_{h|\mathcal{C}}(\cdot|(x,y))$ 不依赖于 y.

*11.7 法则的一致性和几乎必然收敛的实现

首先，考虑用法则序列对取代法则收敛列. 一致性结构(如 2.7 节所定义的一致结构)的量远多于法则空间上的拓扑空间上的积拓扑. 那么将证明对于可分度量空间上的收敛的法则列，存在满足

这些几乎必然收敛法则的随机变量.

11.7.1 定理　对任意可分度量空间(S, d)和S中的法则序列P_n和G_n，当$n\to\infty$时，下列命题等价：

(a)$\beta(P_n, Q_n)\to 0$.

(b)$\rho(P_n, Q_n)\to 0$.

(c)在某个概率空间中存在取值于S的随机变量X_n和Y_n，使得$\forall n$，X_n有法则P_n，Y_n有法则Q_n，且$d(X_n, Y_n)$依概率到0.

(d)将(c)中的"依概率"换为"几乎必然"，结论仍成立.

注：在$\mathcal{R}$中存在一个有界连续函数f，根据定理2.6.4，$f(n)=0$，$f(n+1/n)=1$，$n=1, 2, \cdots$令$P_n=\delta_n$，$Q_n=\delta_{n+(1/n)}$，就可知道对于所有的有界连续函数f，定理11.7.1中的条件(a)和(b)与$\int f\mathrm{d}(P_n-Q_n)\to 0$不等价，虽然根据定理11.3.3知，如果$Q_n$都是相同的，那么它们是等价的.　413

证明　根据推论11.6.5可以得到(b)⇒(a)，由于在11.3.3(c)⇒(d)的证明中证得$\rho\leqslant 2\beta^{1/2}$，因此可证明(a)⇒(b). 显然(d)⇒(c).

(c)⇒(a)：令$\|f\|_{BL}\leqslant 1$，给定$\varepsilon>0$，令A_n是$d(X_n, Y_n)>\varepsilon$的事件. 那么

$$\left|\int f\mathrm{d}(P_n-Q_n)\right| = |E(f(X_n)-f(Y_n))| \leqslant 2Pr(A_n)+\varepsilon.$$

因此，n充分大时有$\beta(P_n, Q_n)\leqslant 3\varepsilon$，即(a)证得.

下面只需证明(b)⇒(d). 定理11.6.3在$\alpha=\beta=\alpha_n=\rho(P_n, Q_n)$和$a=b=a_n+1/n$时，在$S\times S$上给出测度$\mu_n=\eta_n$和$\gamma=\gamma_n$. 令$t_n=\gamma_n(S\times S)$，$T$是可数多个空间$S_n\times S_n$的笛卡儿积，其中$S_n$是$S$的一个副本. 在$I:=[0, 1)$上取勒贝格测度$\lambda$. 令$\Omega:=I\times T$. 对每个$x\in I$，令$Pr_x$是$T$上的概率测度，根据定理8.2.2知，$T$是法则$\mu_n/(1-t_n)$(对于使得$x<1-t_n$的$n$)和$\gamma_n/t_n$(对使得$x\geqslant 1-t_n$的$n$)的笛卡儿积. 令$A$是$\Omega$的任意可测子集. Ω中的法则Pr定义为

$$Pr(A) = \int_I\int_T 1_A(x,y)\,\mathrm{d}Pr_x(y)\,\mathrm{d}x = \int_I Pr_x(A_x)\,\mathrm{d}\lambda(x),$$

其中$A_x:=\{y\in T: \langle x, y\rangle\in A\}$. 为证明$Pr$是明确定义的，首先证明$g_A(x):=Pr_x(A_x)$在$x$处是可测的. 令$\mathcal{M}$是所有集合$C\subset\mathcal{M}$的集族，使得$g_C$是可测的. 令$\mathcal{R}$是$\Omega$上有限维矩形的半环，根据命题8.2.1，在这种情况下，$\mathcal{R}$中的每个集合B都是I中一个博雷尔集的积，对于有限多个n，$\mathcal{R}$中的每个集合B是$S_n\times S_n$中的博雷尔集，对于其他的n，$\mathcal{R}$中的每个集合B也是$S_n\times S_n$的博雷尔集. 对于有限积上的Pr_x，B_x和I中x的博雷尔集，只存在有限多种可能. 它可以得出$\mathcal{R}\subset\mathcal{M}$. 然后根据命题8.2.1知，由$\mathcal{R}$生成的代数$\mathcal{A}$的每一个集合$A$都是$\mathcal{R}$中一个有限不交并$A=\bigcup_i B(i)$. 那么$g_A=\sum_i g_{B(i)}$，所以$g_A$是可测的且$\mathcal{A}\subset\mathcal{M}$. 现在$\Omega$中集合$C(n)$的单调收敛性可以推出函数$g_{C(n)}$的单调收敛性，它保持可测性，因此$\mathcal{M}$是单调类，并且根据定理4.4.2，$\mathcal{M}$包括$\Omega$中的积$\sigma$-代数$\mathcal{S}$，$\mathcal{M}$是由$\mathcal{A}$或$\mathcal{R}$生成的. 显然有$Pr(\Omega)=1$或$Pr(\varnothing)=0$. 由于如果$\mathcal{S}$中的集合$A_j$是不相交的，那么对每个$x\in I$，在$T$中并为$A_x$的集合$(A_j)_x$是不相交的，所以$Pr$是可数可加的，它们的概率$Pr_x$达到$Pr_x(A_x)$，可以运用控制收敛定理或单调收敛定理. 所以$Pr$是$\Omega$上的概率测度.

在I上对Pr_x的边缘关于$\mathrm{d}\lambda(x)$积分可以得到$S_n\times S_n$上的边缘：$\forall n$，$(1-t_n)\mu_n/(1-t_n)+t_n\gamma_n/t_n=\mu_n+\gamma_n$(用1代替0/0)，其在$S$上有边缘$P_n$和$Q_n$. 对于使得$x<1-t_n$的$n$，$S_n\times S_n$上的坐标$(X_n, Y_n)$满足$d(X_n, Y_n)\leqslant\alpha_n+1/n$ a.s.，当n充分大时，这对所有的x成立，因此有　414

$d(X_n, Y_n)\to 0$ a. s.. □

下面将证明如果在定理 11.7.1 中的 Q_n 都是相同的，那么可以取所有的 Y_n 是相同的．这不能作为定理 11.7.1 的推论．其证明方法有所不同并且比较难．

11.7.2 定理 令 S 是任意可分度量空间，$P_n(n=0, 1, \cdots)$ 是 S 中的法则，且当 $n\to\infty$ 时 P_n 收敛到 P_0. 那么在某个概率空间上存在取值于 S 的随机变量 $X_n(n=0, 1, \cdots)$，使得对所有的 n，X_n 有法则 P_n 且 $X_n\to X_0$ a. s..

证明 回忆在定理 11.1.1 中，集合 $A\subset S$ 称为 P_0 连续集，当且仅当 $P_0(\partial A)=0$，其中 ∂A 是 A 的边界，并且对于度量空间 (S, d) 中的集合 A，$\mathrm{diam}(A):=\sup\{d(x, y): x, y\in A\}$. 证明将用到下面的事实．

11.7.3 引理 对任意可分度量空间 (S, d)，P 是 S 中的法则且 $\varepsilon>0$，存在 P 的不相交的博雷尔连续集 A_j，使得 $S=\bigcup_{j\geq 1}A_j$ 和 $\forall j$，$\mathrm{diam}(A_j)$.

证明 $\forall x\in S$ 和 $r>0$，除非 $P\{y: d(x, y)=r\}>0$，否则，球 $B(x, r):=\{y: d(x, y)<r\}$ 是 P 的连续集，它对至多可数多个 r 值是成立的．令 $\{x_j\}_{j\geq 1}$ 在 S 中是稠密的．对每个 j，$B(x_j, r)$ 对于某个 $r=\varepsilon_j(\varepsilon/4<\varepsilon_j<\varepsilon/2)$ 是一个连续集．令 $A_1:=B(x_1, \varepsilon_1)$，对 $j>1$，令

$$A_j := B(x_j,\varepsilon_j)\Big\backslash \bigcup_{i<j} B(x_i,\varepsilon_i).$$

由于连续集形成一个代数(命题 11.1.4)，因此 A_j 是连续集，即引理得证． □

现在继续定理 11.7.2 的证明，令 $P:=P_0$. 令 $\varepsilon:=\varepsilon_m=1/m^2$，$m=1, 2, \cdots$，给定一个 ε，按照引理 11.7.3 中对 P 和 ε 那样，选择连续集 $A_j=A_{jm}$，$j=1, 2, \cdots$. 取 $k(m)$ 足够大，使得 $P\left(\bigcup_{i\leq k(m)}A_{im}\right)>1-\varepsilon$. 对 $\forall i\leq k(m)$，假设 $P(A_{im})>0$(否则给 A_{im} 重新编号或减小 $k(m)$ 的值). 根据定理 11.1.1 知，$\forall j, m$，当 $n\to\infty$ 时，$P_n(A_{jm})\to P(A_{jm})$. 所以对某个足够大的 n_m，$P_n(A_{jm})>p_{jm}:=(1-\varepsilon_m)P(A_{jm})$，其中 $j=1, \cdots, k(m)$，$n\geq n_m$. 用 $\max\{n_i: i\leq m\}$ 替换 n_m，我们可以取非减序列 $\{n_m\}_{m\geq 1}$，也可假设当 $m\to\infty$ 时，$n_m\uparrow\infty$.

令 $n_0:=1$. 对于每个 $n=1, 2, \cdots$，存在唯一的 $m=m(n)$，使得 $n_m\leq n<n_{m+1}$. 对每个博雷尔集 $B\subset S$，令

$$\eta_n(B) := \sum_{1\leq j\leq k(m)} p_{jm}P_n(B\mid A_{jm}),$$

令 $\alpha_n:=P_n-\eta_n$. 根据 p_{jm} 和 n_m 的选择，$P_n\geq\eta_n$，所以 α_n 是 S 上的非负测度．α_n 的全质量是

$$t_n := 1-\sum_{j\leq k(m)}p_{jm} = 1-(1-\varepsilon_m)\sum_{j\leq k(m)}P(A_{jm}) \leq 1-(1-\varepsilon_m)^2 < 2\varepsilon_m.$$

那么

11.7.4
$$\sum_{j\leq k(m)}P(A_{jm}) = (1-t_n)/(1-\varepsilon_m).$$

也有 $t_n\geq 1-(1-\varepsilon_m)=\varepsilon_m=1/m^2>0$.

现在，概率空间 Ω 是 $I\times\prod_{n\geq 0}S_n$，$S_n$ 是 S 的副本．对每个 $t\in I:=(0, 1]$，$x\in S_0$，$n=1, 2, \cdots$ 定义 S_n 上的法则 $\mu_n(x, t)(\cdot)$，如果 $n<n_1$，$\mu_n\equiv P_n$；如果 $m=m(n)\geq 1$，令 $\mu_n(t, x)(\cdot):=P_n(\cdot\mid A_{jm})$；如果 $t\geq\varepsilon_m$，$x\in A_{jm}=A_{jm(n)}$ 且 $j\leq k(m)$. 否则，如果 $x\notin\bigcup_{j\leq k(m)}A_{jm}$ 或 $t<\varepsilon_m$，令 $\mu_n(t, x)(\cdot):=\alpha_n/t_n$. 所以 μ_n 是定义在 S 的博雷尔集上的概率法则．令 λ 是 I 上的勒贝格测度．给

415

定$(t, x)\in I\times S_0$，在 $V:=\prod_{n\geq 1}S_n$ 上令 $\mu_{t,x}$ 是法则 $\mu_n(t, x)(\cdot)(\forall n)$ 的积，根据定理 8.2.2 知，$\mu_{t,x}$ 是存在的．Ω 上的一个概率测度 Pr 定义为 $Pr(B):=\iint_I\int_S\int_V 1_B(x,s,v)\,d\mu_{s,x}(v)\,dP(s)\,dx$. 因此，在 $I\times S_0$ 上的(t, x)将有积法则 $\lambda\times P$. 为证明 Pr 是明确定义的，需要像定理 11.7.1 的证明中那样证明内积分是(s, x)的可测函数．定义有限维矩形，$\mu_n(t, x)$对于有限多个 $n\geq 1$ 的积仅仅存在有限多种可能，每个都出现在 $I\times S_0$ 中的(t, x)的可测集上．因此对于这些矩形可得到所需的可测性．我们可以通过矩形的半环(命题 8.2.1)去构造矩形的有限不相交并的代数，然后根据单调类(定理 4.4.2)得到积 σ-代数(如同定理 11.7.1 的证明). 因此，在 Ω 上可以得到明确定义的法则 Pr.

令 X_n 是 Ω 上的第 n 个坐标，其取值于 S_n. 那么对 $n<n_1$，显然有 $Pr\circ X_n^{-1}=P_n$，因此，假设 $n\geq n_1$ 和 $m=m(n)\geq 1$，$\varepsilon=\varepsilon_m$ 和 $k:=k(m)$. 那么在 I 中关于 t 求积分可以得到 X_n 关于 X_0 的条件分布，如果 $X_0\in A_{jm}$(对某个 $j=1, \cdots, k$)，该条件分布是$(1-\varepsilon)P_n(\cdot\mid A_{jm})+\varepsilon\alpha_n/t_n$；否则，$X_n$ 关于 X_0 的条件分布为 α_n/t_n. 根据(11.7.4)得，$P\left(S\setminus\bigcup_{j\leq k(m)}A_{jm}\right)=(t_n-\varepsilon)/(1-\varepsilon)$. 那么对于 P 关于 X_0 求积分得，

416

$$Pr\circ X_n^{-1}=(t_n-\varepsilon)\alpha_n/(t_n(1-\varepsilon))+\sum_{1\leq j\leq k}P(A_{jm})\{(1-\varepsilon)P_n(\cdot\mid A_{jm})+\varepsilon\alpha_n/t_n\}=\eta_n+c\alpha_n.$$

由于我们得到一个概率法则(可以直接验证)，所以常数 $c=c(n, \varepsilon)$必须是 1. 因此，X_n 的分布是 $\eta_n+\alpha_n=P_n$.

由于 $\sum_m\varepsilon_m=\sum_m 1/m^2<\infty$，由对于 P 的博雷尔–坎泰利引理可以得出对几乎所有的 $X_0\in S_0$，m 足够大时，对某个 $j\leq k(m)$，$X_0\in A_{jm}$. 同样，对所有的 $t\in I$，m 足够大时有 $t>\varepsilon_m$，n 足够大时有 $t>\varepsilon_n$. 当上述情况发生的时候，X_n 关于 X_0 的条件分布集中在 A_{jm} 上，所以 $d(X_n, X_0)<\varepsilon_m$，因此当 $n\to\infty$ 时，$X_n\to X_0$ a. s.. □

定理 11.7.1 和定理 11.7.2 都是对"某个测度空间"而言，有时使用固定的概率空间是有好处的，例如，具有勒比格测度 λ 的单位间隔$[0, 1]$. 下面的定理在$[0, 1]$作为概率空间代替完备可分度量空间时是有用的．

11.7.5 定理 对任意完备可分度量空间(S, d)及 S 上的法则 P，存在一个从$[0, 1]$映射到 S 的博雷尔可测函数 f，使得 $\lambda\circ f^{-1}=P$.

证明 对每个 $m=1, 2, \cdots$，令 $S=\bigcup_n A_{mn}$，其中 A_{mn} 是 P 的不相交的博雷尔连续集，并且根据引理 11.7.3 知，$\operatorname{diam}(A_{mn})\leq 1/m$. 取 $0=t_0\leq t_1\leq\cdots$，使得对 $\forall n$，$P(A_{1n})=t_n-t_{n-1}$. 令正整数的 k-元组$(n(1),\cdots, n(k))$记为(n, k). 令 $B_{(n,k)}:=\bigcap_{1\leq j\leq k}A_{jn(j)}$. 令 $I_{(n,1)}:=[t_{n-1}, t_n)$，对于 $k\geq 2$，$n(k)=1, 2, \cdots$，关于 k 递归地定义左闭右开区间 $I_{(n,k)}$，它构成了 $I_{(n,k-1)}$ 到不相交区间的分解，使得 $\forall(n, k)$，$\lambda(I_{(n,k)})=P(B_{(n,k)})$. 那么当 $P(B_{(n,k)})=0$ 时，$I_{(n,k)}$ 是空的．

定义从$[0, 1]$映射到 S 的函数序列 f_k 如下：当 $I_{(n,k)}\neq\varnothing$ 时，选择 $x(n, k)\in B_{(n,k)}$，$\forall t\in I_{(n,k)}$，令 $f_k(t):=x_{(n,k)}$. 那么当 $k\to\infty$ 时，博雷尔函数 f_k 一致收敛到某个从$[0, 1]$映射到 S 的博雷尔函数 f (根据定理 4.2.2 知，f 是博雷尔可测的). 令 $P_k:=\lambda\circ f_k^{-1}$，那么 $P_k\to\lambda\circ f^{-1}$.

417

现在 $\forall(n, k)$，$P_k(B_{(n,k)})=P(B_{(n,k)})$. 那么对任意博雷尔集 $A\subset S$，有

$$P_k(A) = \sum_{(n,k)} \{\lambda(I_{(n,k)}) : x_{(n,k)} \in A\},$$

(这里的和是关于给定的 k 对所有的 k-元组求和.) $\forall \varepsilon > 0$,

$$P_k(A) = \sum_{(n,k)} \{P(B_{(n,k)}) : x_{(n,k)} \in A\} \leqslant P(A^{\varepsilon+1/k}).$$

其中 A^δ 如同 Prohorov 度量 ρ(11.3 节)定义中的 A^δ. 我们得到 $\rho(P_k, P) \leqslant 1/k$, 根据定理 11.3.3 有 $P_k \to P$, 所以有 $P = \lambda \circ f^{-1}$. □

回忆莱维度量: 对于ℝ上的法则 P 和 Q, 其分布函数分别 F 和 G,

$$\lambda(P,Q) := \inf\{\varepsilon > 0 : \forall x, F(x-\varepsilon) - \varepsilon \leqslant G(x) \leqslant F(x+\varepsilon) + \varepsilon\}$$

那么在ℝ上 λ 可度量化依 L 收敛, 但是在 11.3 节的习题 8 中证得 λ 定义的一致性比(由 ρ 或 β 定义的)Fortet-Prohorov 一致性要弱.

由于在中心极限定理中根据特征函数的逐点收敛性可以证得依 L 收敛(9.5 节和 9.8 节), 但令人惊奇的是即使特征函数的更强(一致)的闭性也不能推出法则的闭性.

11.7.6 命题 *在ℝ上存在法则 P_n 和 Q_n, $\forall n$, 满足 $P_n([1, \infty)) = Q_n((-\infty, -1]) = 1$, 并且其特征函数分别为 f_n 和 g_n, 使得当 $n \to \infty$ 时, $\sup_t |(f_n - g_n)(t)| \to 0$.*

注: 对于这样的 P_n 和 Q_n, 显然满足 $\rho(P_n, Q_n) = 1$, 所以这些法则与 Prohorov 度量相差很远.

证明 令 P_n 在区间 $[1, n]$ 上的密度函数为 $1/(x\log n)$, 在其他地方密度为 0. 由变换 $x \mapsto -x$, 令 Q_n 为 P_n 的像, 所以 $\mathrm{d}(Q_n(x)) = \mathrm{d}(P_n(-x))$. 那么对任意的 t 和 $y := xt$,

$$(f_n - g_n)(t) = (\log n)^{-1} \int_1^n \mathrm{e}^{\mathrm{i}xt} - \mathrm{e}^{-\mathrm{i}xt} \mathrm{d}x/x$$

$$= 2\mathrm{i}(\log n)^{-1} \int_t^{nt} (\sin y)/y \mathrm{d}y.$$

418 现在, 由于从 $m\pi$ 到 $(m+1)\pi$ 的积分符号是交错的并且当 $m \to \infty$ 时其绝对值是递减的, 所以 $\int_0^u (\sin y)/y \mathrm{d}y$ 关于 u 是一致有界的. □

习题

1. 证明: 如果 S 是ℝ上具有通常度量的实直线, 在有勒贝格测度的概率空间 $(0, 1)$ 中定理 11.7.2 成立. [提示: 如同命题 9.1.2 中所定义的, 令 X_n 是 P_n 的分布函数的逆.]
2. 如果 S 是任意完备可分度量空间, 则在有勒贝格测度的概率空间 $(0, 1)$ 中定理 11.7.2 成立. [提示: 根据定理 2.5.7, 完备可分度量空间 S_n 中有积拓扑的可数积像完备可分度量空间一样也是可度量的.]
3. 假设 $\{P_n\}$ 和 $\{Q_n\}$ 是ℝ上一致紧的法则序列, 其特征函数分别是 f_n 和 g_n. 假设对每个 $M < \infty$, 当 $n \to \infty$ 时, $\sup_{|t| \leqslant M} |(f_n - g_n)(t)| \to 0$. 证明: $\rho(P_n, Q_n) \to 0$. [提示: 假设不成立, 取子序列.]
4. 如 (S, d) 和 (T, u) 是度量空间, 从 S 映射到 T 的函数 f 称为等距同构, 当且仅当 $e(f(x), f(y)) = d(x, y)$, $\forall x, y \in S$. 假设 γ 是一个函数, 使得对任意度量空间 (S, d), $\gamma = \gamma_{S,d}$ 是 S 上所有法则集合上的一个度量. 称 γ 是不变定义的(invariantly defined), 如果其满足:
   (a) 对 S 上的任意法则 P 和 Q, 以及从 S 映射到 T 的等距同构 f, 有 $\gamma(P \circ f^{-1}, Q \circ f^{-1}) = \gamma(P, Q)$.
   (b) 如果法则 P 和 Q 对某个可测子集 A, 给出全质量 1, 那么 $\gamma_{A,d}(P, Q) = \gamma_{S,d}(P, Q)$.
   证明: 度量 ρ 和 β 都是不变定义的.
5. 令 F 是 S 上所有函数 f 组成的集合, 其中 $f(x) \equiv \delta(1 - d(x, B)/\varepsilon) \vee 0$, 这里 B 是有界集, 满足 $\mathrm{diam}B := \sup\{d(x, y) : x, y \in B\} < \infty$, 并且这里 $0 < \varepsilon < 1$, $0 < \delta < \min(\varepsilon, 1/\mathrm{diam}B)$.

令 $d_F(P, Q) := \sup\left\{\left|\int f\mathrm{d}(P-Q)\right| : f \in F\right\}$. 在从 S 映射到其他度量空间 T 的等距同构上令 $\bar{d}_F(P, Q) := \sup_{g,T} d_F(P \circ g^{-1}, Q \circ g^{-1})$. 证明：$d_F$ 和 $\bar{d}_F$ 是度量并且在 S 上度量化通常的依 L 收敛 . [提示：见 portmanteau 定理(11.1.1)及其证明，如果在 T 上有 $f \in F$，考虑 $\|f \circ g\|_{BL}$.]

6. 证明：$\bar{d}_F$ 是不变定义的 .
7. 证明：$\bar{d}_F$ 可以定义一个与 β 和 ρ 定义的一致性不同的一致性 . 例如，在实直线 $\mathbb{R}$ 上，如果 P_n 是 1，…，n 上点质量的平均，那么对于 $\bar{d}_F$，P_n 收敛到 0；同理，如果 Q_n 是 -1，…，$-n$ 上点质量的平均，那么对于 $\bar{d}_F$，Q_n 收敛到 0. 所以 $\bar{d}_F(P_n, Q_n) \to 0$，虽然 P_n 和 Q_n 对于 β 或 ρ 不是闭的 .

419

*11.8 Kantorovich-Rubinstein 定理

斯特拉森定理(推论 11.6.4)说明了对于在 Prohorov 度量中距离为 r 的两个法则 P 和 Q，存在具有法则 P 和 Q 的随机变量，其樊畿距离为 r 或者如果空间不是完备的，樊畿距离大于 r. 本节将用其他距离代替 Prohorov 和樊畿距离来证明一个非常类似的事实 .

对于 S 上的任意两个法则 P 和 Q，令 $M(P, Q)$ 表示 $S \times S$ 上边缘为 P 和 Q 的所有法则的集合 . 该集合中的一个元素是积法则 $P \times Q$.

随机变量的距离为 $Ed(X, Y)$. 这个距离不总是有限的，所以我们考虑随机变量有限时的情况 .

11.8.1 命题 令 (S, d) 是一个可分度量空间 . 令 X 是取值于 S 的随机变量 . 那么下列命题等价：

(a) 对某个 $y \in S$，$Ed(X, y) < \infty$.

(b) $Ed(X, z) < \infty$，$\forall z \in S$.

证明 由于 $d(X, z) \leqslant d(X, y) + d(y, z)$，即(a)⇒(b). □

$\mathcal{P}_1(S)$ 表示 S 上所有法则 P 的集合，对某个(因此每个) $z \in S$，$\int d(x, z)\mathrm{d}P(x) < \infty$. 注意到，如果 (S, d) 是有界的，S 上的所有法则都属于 $\mathcal{P}_1(S)$. 例如，如果 $S = \mathbb{R}$ 且 X 是一个实值随机变量，则 $\mathcal{L}(X) \in \mathcal{P}_1(S)$ 当且仅当 $E|X| < \infty$.

对于 $\mathcal{P}_1(S)$ 上的任意两个法则 P 和 Q，Wasserstein 或 Monge-Wasserstein 距离 W 定义为

$$W(P,Q) := \inf\left\{\int d(x,y)\,\mathrm{d}\mu(x,y) : \mu \in M(P,Q)\right\}.$$

由于对任意的 $z \in S$，有 $d(x, y) \leqslant d(x, z) + d(y, z)$，$W(P, Q)$ 在 $\mathcal{P}_1(S)$ 上对任意的 P 和 Q 是有限的 . 这里 W 与运输问题有关 . 令 P 是生产货物点的分布，Q 是消费点的分布 . 令 $d(x, y)$ 是从 x 到 y 运费 . 那么 $W(P, Q)$ 就是把所有货物运输给消费者的最小费用 .

现在回忆在 6.1 节和 11.2 节中介绍的 S 上适当实值函数 f 的利普希茨半范数：

$$\|f\|_L := \sup\{|f(x) - f(y)|/d(x,y) : x \neq y \in S\},$$

420

对 S 上的任意两个法则 P 和 Q，令

$$\gamma(P,Q) := \|P - Q\|_L^* := \sup\left\{\left|\int f\mathrm{d}(P-Q)\right| : \|f\|_L \leqslant 1\right\}.$$

例如，在 $\mathbb{R}$ 上令 $P_n := ((n-1)\delta_0 + \delta_n)/n$，$P := \delta_0$. 那么对于前面定义的依 L 收敛有 $P_n \to P$，因此等价地，对于度量 ρ 和 β 也一样，但是 $\gamma(P_n, P) \equiv 1$，$W(P_n, P) \equiv 1$.

下面是本节的一个主要定理 .

11.8.2 定理(Kantorovich-Rubinstein) 对任意可分度量空间 (S, d) 和 $\mathcal{P}_1(S)$ 上的任意两个法则 P 和 Q，有

$$W(P,Q) = \gamma(P,Q).$$

如果 P 和 Q 是胎紧的，例如，S 是完备的，那么在 $M(P, Q)$ 上存在法则 Pr，使得 $\int d(x, y)\mathrm{d}Pr(x, y) = W(P, Q)$，那么 $W(P, Q)$ 定义中的下确界可以得到.

证明 令 $\mu \in M(P, Q)$ 和 $\|f\|_L \leqslant 1$，那么有

$$\left|\int f\mathrm{d}(P-Q)\right| = \left|\int f(x) - f(y)\mathrm{d}\mu(x,y)\right| \leqslant \int d(x,y)\mathrm{d}\mu(x,y).$$

对所有的 f 取上确界，对所有的 μ 取下确界，给定 $\gamma(P, Q) \leqslant W(P, Q)$. 然后将会用到下面的事实.

11.8.3 引理 W 和 γ 都是 $\mathcal{P}_1(S)$ 上的伪度量.

注：事实上，W 和 γ 都是度量，在证明定理 11.8.2 之后，只需证明两个相等的伪度量之中的一个是度量.

证明 显然，对任意的 P 和 Q，$W(P, Q) = W(Q, P) \geqslant 0$，且 $\gamma(P, Q) = \gamma(Q, P) \geqslant 0$. 事实上，$\gamma$ 显然是伪度量.

令 $\varepsilon > 0$，那么如同在定理 11.6.3 情况 C 中的证明那样，S 是不相交博雷尔集 $S_n(n = 1, 2, \cdots)$ 的可数并，使得 $\forall n$，$x \in S_n$，$y \in S_n$，$d(x, y) < \varepsilon$. 令 $\mu \in M(P, Q)$，使得 $\int d(x, y)\mathrm{d}\mu(x, y) < W(P, Q) + \varepsilon$. 对任意博雷尔集 $A \subset S$，令 $\mu_{mn}(A) := \mu((A \cap S_m) \times S_n)$ 和 $\nu_{mn}(A) := \mu(S_m \times (A \cap S_n))$. 令 $\mu' := \sum_{m,n}(\mu_{mn} \times \nu_{mn})/\mu(S_m \times S_n)$，其中用 0 代替 0/0. 那么 μ' 是 $S \times S$ 上的法则，使得 $\forall m, n$，$\mu'(S_m \times S_n) = \mu(S_m \times S_n)$，这里 μ' 是 $S_m \times S_n$ 上的积测度，$\mu' \in M(P, Q)$，

421

$$\int d(x,y)\mathrm{d}\mu'(x,y) < W(P,Q) + 3\varepsilon.$$

对 S 上的任意其他法则 T，同样有法则 $\nu \in M(Q, T)$，它是一个积测度序列的和，使得

$$\int d(y,z)\mathrm{d}\nu(y,z) < W(Q,T) + 3\varepsilon.$$

现在，给定 $y \in S$，对于 μ'（如前面 10.2 节所定义的），x 关于 y 在每一个博雷尔集 S_m 上的条件分布有可数多个可能的值. 同样对于 ν，z 关于 y 在每一个博雷尔集 B_i 上的条件分布也有可数多个可能的值. 在 $S \times S \times S$ 上定义 (x, y, z) 的法则，使得 (y, z) 有分布 ν，(x, y) 有分布 μ'，这里 y 有分布 Q，且对于每个 y，x 和 z 是条件独立的. 那么 $\langle x, z\rangle$ 关于 y 的可能的条件分布的集合也是可数的，它是个积法则，并且其发生在每一个博雷尔集 $S_m \cap B_i$ 上. 那么法则 M 在 $S \times S \times S$ 是明确定义的，并且其在 S 上有边缘 P，Q 和 T，且满足

$$\int d(x,z)\mathrm{d}M(x,y,z) \leqslant W(P,Q) + W(Q,T) + 6\varepsilon.$$

令 $\varepsilon \downarrow 0$，且在 $M(P, T)$ 上考虑 $\langle x, z\rangle$ 的边缘，有 $W(P, T) \leqslant W(P, Q) + W(Q, T)$，所以 W 是一个伪度量. □

11.8.4 引理 对于任意的 $P \in \mathcal{P}_1(S)$，存在具有有限集 F_n 的法则 P_n，使得 $\forall n$，$P_n(F_n) = 1$，当 $n \to \infty$ 时，$W(P_n, P) \to 0$，$\gamma(P_n, P) \to 0$.

证明 对每个 n，存在不相交的博雷尔集 $S_{n,k}(k = 1, 2, \cdots)$，满足 $\bigcup_k S_{nk} = S$，使得 $\forall x, y \in S_{nk}$，$\forall k$，$d(x, y) < 1/n$. $\forall n$ 和 k，在 S_{nk} 中选择点 x_{nk}，使得 S_{nk} 是非空的. $\forall x \in S_{nk}$，$\forall k$，令

$f_n(x):=x_{nk}$. 同样 $\forall x\in S_{nj}$, $j=1$, $\cdots$, k, 令 $f_{nk}(x):=x_{nj}$; $\forall x\in S(n,j):=S_{nj}$, $j>k$, 令 $f_{nk}(x):=x_{n1}$. (可以假设 S_{n1} 是非空的).

根据控制收敛定理知，当 $k\geqslant k_n$ 充分大时，$\sum_{j>k}\int_{S(n,j)}d(x_{n1},x)\mathrm{d}P(x)<1/n$，那么 $\int d(f_{nk}(x),x)\mathrm{d}P(x)<2/n$. 令 $P_n:=P\circ f_{nk}^{-1}$. 那么对于 $S\times S$ 上的 $\langle x,f_{nk}(x)\rangle$ 的分布 μ_n，其中 x 在 X 上有分布 P，我们有 $\mu_n\in M(P,P_n)$ 和 $\int d(x,y)\mathrm{d}\mu_n(x,y)<2/n$，所以 $W(P,P_n)\leqslant 2/n$. 由于如同在定理 11.8.2 证明的开始所述 $\gamma\leqslant W$，即引理 11.8.4 得证. □

现在为证明定理 11.8.2，首先假设 S 是紧的. 对于 $S\times S$ 上的任意连续实值函数 h 和 S 上的法则 P 和 Q，令

$$m_{P,Q}(h):=\sup\left\{\int f\mathrm{d}P+\int g\mathrm{d}Q: f(x)+g(y)<h(x,y),\forall x,y\right\}.$$

422

令 $C(S\times S)$ 是 $S\times S$ 上所有通常积拓扑的连续实值函数的空间.

11.8.5 引理　对于任意紧的度量空间 (S,d)，及 S 上的法则 P 和 Q，$h\in C(S\times S)$，

$$m_{P,Q}(h)=\inf\left\{\int h\mathrm{d}\mu:\mu\in M(P,Q)\right\}.$$

证明　对任意的 $\mu\in M(P,Q)$，显然，$m_{P,Q}(h)\leqslant\int h\mathrm{d}\mu$. 因此在状态方程中，"$\leqslant$"成立. 对于逆不等式，对于 S 中的任意两个连续实值函数 f 和 g，令 L 是所有函数 $\varphi(x,y)=f(x)+g(y)$ 的线性空间，且令

$$\gamma(\varphi):=\int f\mathrm{d}P+\int g\mathrm{d}Q.$$

所以 r 是明确定义的，因为如果 $f(x)+g(y)\equiv k(x)+j(y)$，那么 $f(x)-k(x)\equiv j(y)-g(y)$，其必须是某个常数 c，所以有 $k=f-c$ 和 $j=g+c$，可以得出 $\int f-k\mathrm{d}P+\int g-j\mathrm{d}Q=0$.

对于任意的 $h\in C(S\times S)$，令

$$U:=U_h:=\{f\in C(S\times S):f(x,y)<h(x,y),\forall x,y\}.$$

那么 U 对于上确界范数是开的凸集(由于 S 是紧的). 由于对 $\forall x,y$, $f(x)+g(y)<h(x,y)$ 可以推出 $r(\varphi)\leqslant\sup(f)+\sup(g)\leqslant\sup(h)<+\infty$，所以 r 在 L 上是一个线性范数，不恒等于 0，并且其在非空凸集 $U\cap L$ 上是有界的. 因此由哈恩-巴拿赫定理(6.2.11)知，γ 可以扩张为 $C(S\times S)$ 上的线性形式 ρ，满足 $\sup_{u\in U}\rho(u)=\sup_{v\in U\cap L}r(v)$.

$\forall x,y$，假设 $u\in C(S\times S)$，$u(x,y)\geqslant 0$. 那么 $\forall C\geqslant 0$，$h-1-cu\in U$，所以当 $c\to+\infty$ 时，$\rho(h-1-cu)$ 有上界，这蕴涵着 $\rho(u)\geqslant 0$. 同样对任意的 $f\in C(S\times S)$，$|\rho(f)|\leqslant\rho(1)\sup|f|$，所以 $\rho(\cdot)$ 是有界的. 根据 Riesz 表示定理(7.4.1)知，在 $S\times S$ 上存在一个非负有限测度 ρ，使得 $\forall f\in C(S\times S)$，$\rho(f)=\int f\mathrm{d}\rho$. 由于在 L 上 $\rho=r$，故对任意的 f 和 $g\in C(S)$，$\int f(x)\mathrm{d}\rho(x,y)=\int f\mathrm{d}P$ 和 $\int g(y)\mathrm{d}\rho(x,y)=\int g\mathrm{d}Q$. 那么由于 S 是紧度量空间，根据 Riesz 表示定理(7.4.1)的唯一性和像测度定理(4.1.11)知，$\rho\in M(P,Q)$. 现在有

$$m_{P,Q}(h)=\sup_{u\in U\cap L}r(u)=\sup_{U}\rho=\int h\mathrm{d}\rho,$$

所以 $m_{P,Q}(h)\geqslant\inf\{h\mathrm{d}\mu:\mu\in M(P,Q)\}$，即引理 11.8.5 得证. □

423

11.8.6 引理 如果 S 是一个紧度量空间，$h\in C(S\times S)$，下列命题是等价的：

(a)存在一个集合 $J\subset C(S)$，使得对于 S 上任意法则 P 和 Q，有

$$m_{P,Q}(h) = \sup_{j\in J}\left|\int j\mathrm{d}(P-Q)\right|.$$

(b)h 是 S 上的伪度量. 如果 h 是一个伪度量，我们可以取

$$J := J_h := \{j\in C(S): |j(x)-j(y)|\leqslant h(x,y), \forall x,y\}.$$

证明 如果 J 存在，对于任意的 x 和 y，取 $P=\delta_x$ 和 $Q=\delta_y$. 那么 $M(P, Q)$仅包含 $\delta_{\langle x,y\rangle}$，所以根据引理 11.8.5，

$$h(x,y) = \sup_{j\in J}|j(x)-j(y)|,$$

它是一个伪度量.

反之，如果 h 是一个伪度量，$\forall x, y, f(x)+g(y)<h(x, y)$，令 $j(x):=\inf_y(h(x, y)-g(y))$. 那么 $f\leqslant j\leqslant -g$，并且对于所有的 x 和 x'，有

$$j(x)-j(x')\leqslant \sup_y(h(x, y)-h(x', y))\leqslant h(x, x'),$$

所以有 $j\in J_h$ 和 $\int f\mathrm{d}P+\int g\mathrm{d}Q\leqslant\int j\mathrm{d}(P-Q)$，因此有

$$m_{P,Q}(h)\leqslant \sup_{j\in J}\left|\int j\mathrm{d}(P-Q)\right|.$$

由于对任意的 $j\in J$，在 $m_{P,Q}$的定义中可以令 $f=j$，$g=-j-\delta$，$\delta\downarrow 0$，逆不等式总是成立的. □

现在对于一个紧度量空间(S, d)，在引理 11.8.5 中令 $h=d$，有 $m_{P,Q}(d)=W(P, Q)$. 那么根据引理 11.8.6，$W(P, Q)=\|P-Q\|_L^*$，由于 S 是紧的，即可证明定理 11.8.2. 在更一般的情况下，给定 S 上的任意两个法则 P 和 Q，根据引理 11.8.4，取具有有限个支撑的 P_n 收敛到 P，同理，取 Q_n 收敛到 Q. 对于每个 n，$W(P_n, Q_n)=\|P_n-Q_n\|_L^*$. 由伪度量的性质(引理 11.8.3)和引理 11.8.4 的收敛性得，$W(P, Q)=\|P-Q\|_L^*$，即证得定理 11.8.2. □

424

习题

1. 证明：对于在$\mathbb{R}^1$ 上分别具有分布函数 F 和 G 的法则 P 和 Q，$W(P,Q)=\int_{-\infty}^{\infty}|F-G|(x)\mathrm{d}x$. [提示：令 $h:=21_{\{F>G\}}-1$，使得 $h=\begin{cases}1, & 若 F>G\\ -1, & 若 F\leqslant G\end{cases}$. 令 H 是关于 h 的不定积分，有 $\|H\|_L=1$. 证明对任意满足 $\|J\|_L\leqslant 1$ 的函数f，

$$\left|\int J d(P-Q)\right| = \left|\int (F-G)(x)J'(x)dx\right|\leqslant\int|F-G|(x)\mathrm{d}x$$

且对于 $J=H$，这个界是可得的.]

2. 对于分布函数 F 的逆 X_F(如命题 9.1.2 中所定义的)，证明：$\int|F-G|(x)\mathrm{d}x=\int_0^1|X_F-X_G|(t)\mathrm{d}t$. [提示：这是两个图之间的面积.(从而根据在一个概率空间(事实上，具有勒贝格测度的单位区间)上具有这些法则的随机变量知，$\mathbb{R}^1$ 上所有法则之间的 Monge-Wasserstein 距离可以同时得到).]

3. 证明：在$\mathbb{R}^2$ 上存在 3 个法则 α，β 和 γ，使得在$\mathbb{R}^6$ 上不存在法则 P，在$\mathbb{R}^2$ 具有坐标 x，y，和 z，x，y，和 z 的法则分别为 α，β 和 γ 且 $E|x-y|=W(\alpha, \beta)$，$E|y-z|=W(\beta, \gamma)$和 $E|x-z|=W(\alpha, \gamma)$. [提示：令 a，b 和 c 是边长为 1 的等边三角形的顶点. 令 $\alpha:=(\delta_a+\delta_b)/2$，$\beta=(\delta_a+\delta_c)/2$ 和 $\gamma=(\delta_b+\delta_c)/2$.(所以如同习题 2 中在$\mathbb{R}^1$ 上那样，在$\mathbb{R}^2$ 上对于所有法则的随机变量的 Monge-Wasserstein 距离不可能同时达到).]

4. 令 P 是一个可分赋范向量空间$(S, \|\cdot\|)(x\in S)$上的一个法则，P_x 是 P 平移 x 单位得到的，所以对任意博雷尔集 A，$P_x(A):=P(A-x)$，其中 $A-x:=\{a-x: a\in A\}$. 证明 $W(P, P_x)=\|x\|$.
5. 证明：W 是$\mathcal{P}_1(S)$上的一个度量. [提示：考虑度量 β.]
6. 对于可分度量空间(S, d)和 $1\leqslant r<\infty$，令$\mathcal{P}_r(S)$是 S 上所有法则 P 的集合，使得对某个 $y\in S$，$\int d(x, y)^r dP(x)<\infty$. 证明：对于所有的 $y\in S$，结论同样成立.
7. $\forall 1\leqslant r<\infty$，$P$ 和 $Q\in\mathcal{P}_r(S)$，令 $W_r(P, Q):=(\inf\{Ed(X, Y)=P, \mathcal{L}(Y)=Q\})^{1/r}$. 证明：对于$\mathcal{P}_r(S)$上的任意法则 P，存在有限集 $F_n\subset S$ 和法则 P_n，满足 $P_n(F_n)=1$，使得当 $n\to\infty$ 时，$W_r(P_n, P)\to 0$，如同引理 11.8.4 中 $r=1$ 的情况.
8. 令 S 是一个波兰空间，令 μ 和 ν 是 $S\times S$ 上的法则. 对于$\langle x, y\rangle\in S\times S$，令 $p_1(x, y):=x$，$p_2(x, y):=y$. 假设$\mu\circ p_2^{-1}=\nu\circ p_1^{-1}$ 是 S 上的法则，证明：存在 $S\times S\times S$ 上的法则 τ，对 $f(x, y, z):=\langle x, y\rangle$ 和 $g(x, y, z):=\langle y, z\rangle$，我们有 $\tau\circ f^{-1}=\mu$ 和 $\tau\circ g^{-1}=\nu$. [提示：见引理 11.8.3 的证明.] 425
9. 证明：对任意的波兰空间，W_r 是$\mathcal{P}_r(S)$上的一个度量. [提示：运用习题 8 和 10.2 节.]

*11.9 U-统计量

对于独立同分布变量的对称函数的平均变量，中心极限定理成立. 令 X_1，X_2，…是具有法则 Q 的独立同分布变量，取值于某个可测空间$(S, \mathcal{A})$，其中通常 $S=\mathbb{R}$ 且有博雷尔 σ-代数. 令 f 是 m 个 S 的笛卡儿积 S^m 上的一个可测函数，因此可以求 $f(X_1, \cdots, X_m)$的值. 假设 f 和 Q 满足 $E|f(X_1, \cdots, X_m)|<\infty$. 那么集合 $g(Q):=Ef(X_1, \cdots, X_m)$. 现在对于某个 $n\geqslant m$，给定 X_1，…，X_n. 在统计中，我们假设只给定 X_1，…，X_n 的数据或观测值. 除了由 X_1，…，X_n 提供的关于 Q 的信息和其他一些正则性条件外，Q 是未知的，在这种情况下，$g(Q)$是有限的. 一个**统计量**(statistic)是 X_1，…，X_n 的任意可测函数 T_n. 统计量 T_n 称为是函数 $g(Q_n)$的**无偏估计**(unbiased estimator)，当且仅当在 X_1，…，X_n 是独立同分布 Q 时，$ET_n(X_1, \cdots, X_n)=g(Q)$. 统计量序列$\{T_n\}$(其中 $T_n=T_n(X_1, \cdots, X_n)$)称为是 $g(Q)$估计的**相合序列**(consistent sequence)，当且仅当 X_1，X_2…是独立同分布变量且其分布是 Q，当 $n\to\infty$ 时，T_n 收敛到 $g(Q)$.

如果 $g(Q)=Ef(X_1, \cdots, X_m)$，得到 $g(Q)$的无偏估计的相合序列的一种方法是设 $Y_1=f(X_1, \cdots, X_m)$，$Y_2=f(X_{m+1}, \cdots, X_{2m})$，等等，所以 Y_1，Y_2，…是独立同分布变量且 $EY_i=g(Q)$，根据强大数定律知，当 $k\to\infty$ 时，$Y=\overline{Y}^{(k)}=(Y_1+\cdots+Y_k)/k\to g(Q)$ a. s.. 对 $km\leqslant n<(k+1)m$，可以令 $T_n=\overline{Y}^{(k)}$. 这种方法对于 n 个 X_j 的值给出了 n/m 个 Y_i 的值. 这个方法是相对无效的. U-统计量更多地利用 X_1，…，X_n 的信息以对于给定的 n 得到 $g(Q)$的更好的逼近.

m 个变量 x_1，…，x_m 的函数 f 称为**对称的**，如果对每个$\{1, \cdots, m\}$的排列 π，有 $f(x_1, \cdots, x_m)\equiv f(x_{\pi(1)}, \cdots, x_{\pi(m)})$，换句话说，任意的从$\{1, \cdots, m\}$映上到自身的 1-1 函数 π. 存在 $m!$ 个这样的排列. 对于 $m=2$，f 是对称的当且仅当 $\forall x, y$，$f(x, y)=f(y, x)$. 对于 m 个变量的任意函数 f，**对称化**(symmetrication)由下式定义：

$$f^{(s)}(x_1,\cdots,x_m):=m!^{-1}\sum_{\pi}f(x_{\pi(1)},\cdots,x_{\pi(m)}),$$ 426

这里求和是对所有 $m!$ 个$\{1, \cdots, m\}$的排列 π 来进行的. 那么显然 $f^{(s)}$是对称的，且 f 是对称的当且仅当 $f\equiv f_s$. 例如，如果 $f(x, y)\equiv x+3y$，那么 $f^{(s)}(x, y)=2x+2y$.

给定如上所述的 S 和 f，$n\geqslant m$，第 n 个 U-统计量 $U_n=U_n(f)$定义为

$$U_n := \frac{(n-m)!}{n!} \sum_{i_1,\cdots,i_m \neq}^{n} f(X_{i_1},\cdots,X_{i_m}),$$

这里求和号的意思是对从 1 到 n 的相异整数 i_1，…，i_m 的所有有序 m-元组求和．准确地说，有 $n!/(n-m)!$ 个这样的有序 m-元组（有 n 种方法来选择 i_1，那么有 $n-1$ 种方法选择 i_2，…有 $n-m+1$ 种方法选择 i_m）．所以 U_n 是项的平均，其中每一项都是 f 在具有分布 Q 的 m 个独立同分布变量上的估计，因此，每一项都有期望 $g(Q)$，且 $\forall n$，$EU_n=g(Q)$．

对于每一个从 1 到 n 的相异整数的集合 $\{i_1, \cdots, i_m\}$，存在 $m!$ 个可能的排列，每一个排列都在平均中出现．所以 U_n 没有变化，f 可以用它的对称化 $f^{(s)}$ 来代替，因此 U_n 可以写成更少项的平均：

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1\le i_1<i_2<\cdots<i_m\le n} f^{(s)}(X_{i_1},\cdots,X_{i_m}).$$

例：对于 $m=1$，一个 U-统计量 U_n 是 n 个独立同分布变量 $f(X_i)$ 的平均．例如，对于 $f(x)\equiv x$，样本均值 $\bar{X}$. 然后，对于 $n\geqslant 2$，令 s^2 是样本方差：

$$s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

令 $f(x, y):=(x-y)^2/2$. 那么对于这个 f，将证明 s^2 是 U_n 统计量．考虑两边的公因式，这等价于 $\left(\sum_i X_i^2\right)-n\bar{X}^2=n^{-1}\sum_{i<j}(X_i-X_j)^2$. 要验证这一点，注意到对于 $k=1$，…，n，在每一边 X_k^2 的系数都是 $1-1/n$，且 $i<j$ 时，X_i 和 X_j 的系数都是 $-2/n$.

本节剩下的部分中假设 f 是对称的，即 $f=f^{(s)}$.

如果对于某个 k，h 是定义在 S^k 上的随机变量，令 $E_Q h:=Eh(X_1, \cdots, X_k)$，其中 X_1，…，X_k 独立同分布且有法则 Q. 那么 $g(Q)=E_Q f$. 现在假设 $E_Q f^2<\infty$. 为了去证明在 $n\to\infty$ 时它们的渐近正态性，目前的问题就是去估计和分析 U-统计量 U_n 的方差．

427

对于任意的可积实值随机变量 V 和取值于某可测空间 $(M, \mathcal{B})$ 的随机变量 X. 条件期望 $E(V|X)$ 定义为一个随机变量 $v(X)$，$v(X)$ 是 X 的可测函数且对任意的 $A\in\mathcal{B}$，$E1_A(X)V=E(1_A v)(X)$. 那么在最小 σ-代数上 $v(X)$ 是 V 的条件期望，在这个 δ-代数上 X 是可测的（由定理 4.2.8 知）.

11.9.1 命题　令 X 和 Y 是分别取值于 A 和 B 的独立随机变量，其中 $(A, \mathcal{S})$ 和 $(B, \mathcal{T})$ 是可测空间．令 f 在 $A\times B$ 上是实值可测的且 $E|f(X, Y)|<\infty$. 令 Y 在 B 上有法则 μ. 那么

(a) $f_c(X):=E(f(X, Y)\mid X)=\int f(X, y)\,\mathrm{d}\mu(y)$.

(b) 如果 X，Y 和 Z 是联合独立的，Z 在 B 上取值且有同样的法则 μ，$Ef^2(X, Y)<\infty$，那么 $E(f(X, Y)f(X, Z))=Ef_c^2(X)$.

注：在下面的情况下将会用到这一事实．$A=S^k$，$B=S^{m-k}$，$X=(X_{i(1)}, \cdots, X_{i(k)})$，以及如同 U-统计量的定义中 $Y=(X_{i(k+1)}, \cdots, X_{i(m)})$，同样有 $Z=(X_{j(k+1)}, \cdots, X_{j(m)})$，对 $r=1$，…，k，$i(r)=j(r)$，且集合 $\{i(k+1), \cdots, i(m)\}$ 和 $\{j(k+1), \cdots, j(m)\}$ 是不相交的．这里指标 $i(r)$ $(r=1, \cdots, m)$ 的次序不必是递增的，同样 $j(1)$，…，$j(m)$ 的次序也不必是递增的．

证明　(a) 可直接由定义和 Tonelli-Fubini 定理得到．对于 (b)，记

$$E(f(X,Y)f(X,Z)) = E(\{f(X,Y) - f_c(X) + f_c(X)\}\{f(X,Z) - f_c(X) + f_c(X)\}).$$

那么显然$E(\{f(X, Y) - f_c(X)\}\{-f_c(X) + f_c(X)\}) = 0$. 由于如果先对 Z 求积分，在(a)中$f(X, Z)$可以由$f_c(X)$取代，并且根据条件詹森不等式(10.2.7)得，$Ef_c^2(X) < \infty$，根据 Tonelli-Fubini 定理得，$E(f_c(X)\{f(X, Z) - f_c(X)\}) = 0$. 同理，首先对 γ 求积分有 $E(\{f(X, Y) - f_c(X)\}f(X, Z)) = 0$，那么(b)得证．

□

对于任意函数 $h(x_1, \cdots, x_m)$，当 $X_1, \cdots, X_m$ 是 i. i. d. (Q)时，令 $V_Q(h)$是 $h(X_1, \cdots, X_m)$的方差．那么当 $E_Q f^2 < \infty$ 时，$V_Q(f)$有定义且是有限的．

428

下面的引理将涉及超几何概率：对于任意的非负整数 k, m, n 和 N，令 $h(k, m, n, N)$是一个给定的有 N 个元素的集合 F 的概率，且 F 的子集 G 有 n 个元素，如果随机选择另一个有 m 个元素的子集 H，且对于每一个可能的 H 都有相同的概率，那么交集 $G\cap H$ 恰有 k 个元素．令 G^c 表示补集 $F\setminus G$. 四个集合 $G\cap H$, $G\cap H^c$, $G^c\cap H$ 和 $G^c\cap H^c$ 中每一个的元素个数都可写为如下的一个 2×2 列联表：

	G	G^c	总数
H	k	$m-k$	m
H^c	$n-k$	$N-m-n+k$	$N-m$
总数	n	$N-n$	N

根据计数可能集合的个数，一个估计为

$$h(k,m,n,N) = \frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}} = \frac{n!(N-n)!m!(N-m)!}{k!(n-k)!(m-k)!(N-m-n+k)!N!},$$

这里除了 $0\leqslant k\leqslant m\leqslant N$, $k\leqslant n\leqslant N$，和 $N-m-n+k\geqslant 0$ 的情况外，$h(k, m, n, N) = 0$.

令 $f_k(X_1, \cdots, X_k) := E(f(X_1, \cdots, X_m) \mid X_1, \cdots, X_k)$，所以 f_k 是命题 11.9.1 中的函数 f_c，且有 $X = (X_1, \cdots, X_k)$和 $Y = (X_{k+1}, \cdots, X_m)$. 这是 $V_Q(U_n)$方差分析的一个步骤．

11.9.2 引理　$V_Q(U_n) = \sum_{k=1}^{m} h(k,m,m,n)V_Q(f_k)$.

证明　U_n 的方差是$f_G := f(\{X_i\}_{i\in G})$和$f_H := f(\{X_j\}_{j\in H})$协方差的平均，其中 G 和 H 相互独立地在$\{1, \cdots, n\}$的所有子集构成的集合上取值，且它们的元素均为 m 个．由于对所有的 G, $Ef_G = E_Q f$，故上面的事实成立．如果 G 和 H 共同拥有 $k(k=1, \cdots, m)$个元素，根据命题 11.9.1(b)知，协方差是 $V_Q(f_k)$. 如果 G 和 H 的交为空，那么f_G 和f_H 是独立的并且它们的协方差是 0. 因此由超几何概率的定义可证得结论成立．

□

回忆对任意的随机变量 Y, $\sigma(Y) := (\mathrm{Var}Y)^{1/2}$. 这里有一个关于 U-统计量渐近正态性的定理．

429

11.9.3 定理　当 $f(X_1, \cdots, X_m)$是一个随机变量，且 $E_Q f^2 < \infty$，那么当 $n\to\infty$ 时，有

(a) $Z_n := n^{1/2}(U_n - g(Q)) \underset{\mathcal{L}}{\to} N(0, m^2 V_Q(f_1))$，这里 $N(0, 0) := \delta_0$.

(b)如果 $V_Q(f_1) > 0$，那么$(U_n - g(Q))/\sigma(U_n) \underset{\mathcal{L}}{\to} N(0, 1)$.

证明　首先，根据假设，当 $k = m$ 时，$E_Q f_k^2$和 $V_Q(f_k)$是有限的，又由于条件詹森不等式(10.2.7)知，对于 $k = 1, \cdots, m-1$ 有 $E_Q f_k^2$和 $V_Q(f_k)$也是有限的．

令 $Y_n := mn^{-1/2}\sum_{1\le j\le n}(f_1(X_j)-g(Q))$. 那么 Y_n 是均值为 0、方差为 $m^2V_Q(f_1)$ 的 n 个独立同分布变量和的 $n^{-1/2}$ 倍. 所以根据中心极限定理知，当 $n\to\infty$ 时，$Y_n \xrightarrow[\mathcal{L}]{} N(0,\ m^2V_Q(f_1))$.

下面的引理将要用来证明 $S=\mathbb{R}$，$V_n=Z_n$ 和 $W_n=Y_n$ 的情形. 它是依概率收敛蕴涵依 L 收敛(命题 9.3.5)的一个小的推广.

11.9.4 引理 令$(S,\ d)$是任意可分度量空间，令 μ 是 S 上的法则且 V_n 和 W_n 是取值于 S 上的随机变量，且当 $n\to\infty$ 时，有 $W_n\xrightarrow[\mathcal{L}]{}\mu$ 和 $d(V_n,\ W_n)$依概率收敛于 0. 那么当 $n\to\infty$ 时，$V_n\xrightarrow[\mathcal{L}]{}\mu$.

证明 根据定理 11.7.1，对于 Prohorov 度量 ρ，当 $n\to\infty$ 时，$\rho(\mathcal{L}(V_n),\ \mathcal{L}(W_n))\to 0$. 然后根据定理 11.3.3 得，$\rho(\mathcal{L}(W_n),\ \mu)\to 0$，所以 $\rho(\mathcal{L}(V_n),\ \mu)\to 0$，由此可以推出 $V_n\xrightarrow[\mathcal{L}]{}\mu$. □

由于 $E(U_n-Y_n)^2\to 0$ 蕴涵着 $|U_n-Y_n|$依概率收敛到 0，下一个引理将完成定理 11.9.3(a)的证明.(回顾 $a_n=O(b_n)$意味着 a_n/b_n 在 $n\to\infty$ 时是有界的.)

11.9.5 引理 对每个 m，有如同定理 11.9.3 中的 Z_n 和其证明中的 Y_n，

$$E((Z_n-Y_n)^2) = O(1/n)\to 0, n\to\infty.$$

证明 显然有

$$E((Z_n-Y_n)^2) = EZ_n^2+EY_n^2-2E(Z_nY_n).$$

对 $\forall n$，$EY_n^2=m^2\mathrm{Var}(f_1)$. 由引理 11.9.2 有，

430

$$EZ_n^2 = n\mathrm{Var}(U_n) = \sum_{1\le k\le m} nh(k,m,m,n)V_Q(f_k).$$

现在，

$$h(1,m,m,n) = \binom{m}{1}\binom{n-m}{m-1}\Big/\binom{n}{m} = \frac{m^2(n-m)!^2}{(n-2m+1)!n!}$$

$$= \frac{m^2(n-m)(n-m-1)\cdots(n-2m+2)}{n(n-1)\cdots(n-m+1)}$$

并且当 $n\to\infty$ 时，对 $j=0,\ 1,\ \cdots,\ m-2$，$(n-m-j)/(n-1-j)=1+O(1/n)$，所以有

$$nh(1,m,m,n)V_Q(f_1) = m^2V_Q(f_1)+O(1/n).$$

对于 $k\geqslant 2$，

$$nh(k,m,m,n) = n\binom{m}{k}\binom{n-m}{m-k}\Big/\binom{n}{m} = \frac{nm!^2(n-m)!^2}{k!(m-k)!^2(n-2m+k)!n!}.$$

这就等价于因式 $C(m,\ k)$不依赖于 n 倍的

$$\frac{(n-m)(n-m-1)\cdots(n-2m+k+1)}{(n-1)(n-2)\cdots(n-m+1)},$$

其中在分子上有 $m-k$ 个因子，在分母上有 $m-1$ 个因子. 所以当 $n\to\infty$ 时，这个量是 $O(n^{1-k})$阶的. 由于 m 是固定的且是有限的，所以当 $n\to\infty$ 时，$EZ_n^2=m^2V_Q(f_1)+O(1/n)$.

下面将证明 Y_n 是 Z_n 到平方可积随机变量的线性子空间的正交投影，其中的平方可积随积变量是由随机变量 $h(X_j)(j=1,\ \cdots,\ n)$生成的，它们都是一元函数. 由于 Y_n 是这个形式的，下面只需证明 Z_n-Y_n 与 Y_n 是正交的，也就是说，$EY_n(Z_n-Y_n)=0$ 或 $E(Y_nZ_n)=EY_n^2$，为此，我们有

$$E(Y_nZ_n) = m\binom{n}{m}^{-1}\sum_{i=1}^{n}\sum_{1\le i_1<\cdots<i_m\le n}E(\{f_1(X_i)-g(Q)\}$$

$$\times \{f(X_{i_1},\cdots,X_{i_m}) - g(Q)\}),$$

除非 i 是 i_j 中的一个，否则由独立性知，直和项是 0，在这种情况下，该项是 $V_Q(f_1)$，如同命题 11.9.1 的证明，有 $f=f-f_1+f_1$. 选择 $i_1<\cdots<i_m$，其中一个是 i，对于每一个 i，有 $\binom{n-1}{m-1}$ 种选法，因此，外面的 $\sum_{i=1}^{n}$ 的作用是用 n 乘以这个表达式. 由于

$$mn\binom{n-1}{m-1}\Big/\binom{n}{m} = m^2,$$

故 $E(Y_nZ_n)=m^2V_Q(f_1)=EY_n^2$. 从而当 $n\to\infty$ 时，

$$E((Y_n-Z_n)^2) = \mathrm{Var}(Z_n) - \mathrm{Var}(Y_n) = O(1/n).$$

即证明了引理 11.9.5、引理 11.9.4 和定理 11.9.3(a). 现在如果 $V_Q(f_1)>0$，则有

431

$$\mathrm{Var}(Z_n) = \mathrm{Var}(Y_n) + O(1/n) = m^2V_Q(f_1) + o(1/n), n\to\infty,$$
$$(U_n - EU_n)/\sigma(U_n) = Z_n/\sigma(Z_n),$$

即(b)得证. □

习题

1. 令 X_1, $X_2\cdots$是方差有限的独立同分布变量，且均值 μ 未知. 令 $Y=c_1X_1+\cdots+c_nX_n$ 是 μ 的无偏估计. 证明：$E(Y-\mu)^2$ 对于 $Y=\bar{X}(c_j\equiv 1/n)$是最小的，且是一个 U-统计量. [提示：首先证明 $\sum c_j=1$，然后运用拉格朗日乘数法，设 $E(Y-\mu)^2+\lambda\left(\sum_{j=1}^{n}c_j-1\right)$对于每个 c_j 的偏导数等于 0.]
2. 通常的样本方差 s^2 是一个 2 阶的 U-统计量. 对于 $n=4$，$(X_1+X_2-X_3-X_4)^2/4$ 是方差的另一个可能的估计. 证明：它是 X_1 方差的无偏估计，但是至少在独立同分布的 X_i 有正态分布时它的方差比 s^2 的方差要大. [提示：如果 X 有一个标准正态分布 $N(0, 1)$，那么 $EX^3=0$，$EX^4=3$.]
3. 假设 Y 服从泊松分布，即 $P(Y=k)=\mathrm{e}^{-\lambda}\lambda^k/k!$，$k=0, 1, \cdots$假设仅仅当 $Y\geqslant 1$ 时，Y 是可观测的，我们没有 λ 的其他信息. 现在的问题是估计 $\mathrm{e}^{-\lambda}$. 令 V 是 $\mathrm{e}^{-\lambda}$的一个无偏估计，所以 $V(k)$定义为 $\sum_{k\geqslant 1}\mathrm{e}^{-\lambda}\lambda^kV(k)/(k!\,(1-\mathrm{e}^{-\lambda}))=\mathrm{e}^{-\lambda}$，其中 $k=0, 1, \cdots$，$\forall\lambda>0$. 求 $V(k)$，并讨论其性质及估计的有效性.
4. 对于给定的 $n\geqslant 2$，考虑二项分布族，在 n 次独立试验中成功的次数为 K，每次试验中成功的概率为 p. 那么 K/n 是 p 的通常的无偏估计. 证明：对于 $n=2$，存在 p^2 的唯一的无偏估计. 对于哪些 K 的值，它的值是可标识的？
5. 令 Q 贯穿所有法则的集族，$E_Qf^2<\infty$. 对于 U-统计量 U_n，证明：
   (a) U_n 是 $g(Q)$的唯一无偏估计 T，其对于 X_1，…，$X_n(n\geqslant m)$是对称的. [提示：如果 $V=T-U_n$，则 $E_QV=0$，如果 $Q\{x_i\}=p_i$，$i=1$，…，r，那么 E_QV 是一个对称多项式 $S(p_1, \cdots, p_r)$，且是 n 阶齐次的，当 $p_i\geqslant 0$ 时，$E_QV=0$. (限制 $\sum_i p_i=1$ 可以去掉). 通过对 r 用归纳法证明 $S\equiv 0$，故 $V\equiv 0$.]
   (b) U_n 是具有最小方差 $g(Q)$ 的无偏估计. [提示：对于一个具有对称化 W 的无偏估计 T，根据(a)有 $W=U_n$.]

432

6. 对于独立同分布变量 X_1，X_2，…，令$\mathcal{A}_n$是在其上 X_1，…，X_n 可测的最小 σ-代数，令$\mathcal{B}_n$是在其上 X_{n+1}，X_{n+2}，…可测的最小 σ-代数. 令$\mathcal{S}_n$是$\mathcal{A}_n$中集合的 σ-代数，其示性函数是 X_1，…，X_n 的对称函数. 令$\mathcal{C}_n$是由$\mathcal{S}_n$和$\mathcal{B}_n$生成的 σ-代数. 证明：如上面所述，对由随机变量$f(X_1, \cdots, X_m)$生成的 U-统计量序列$\{U_n, n\geqslant m\}$，$\{U_n, \mathcal{C}_n\}_{n\geqslant m}$是一个逆鞅，其中 f 是对称的且 $E_Q|f|<\infty$.

注释

11.1 节 至少在中心极限定理的特殊情况下，$\mathbb{R}$ 上的 L 收敛可以追溯到棣莫弗和拉普拉斯(见 9.4 节的注释). 在$\mathbb{R}$ 上更一般的情况，根据 Hildebrandt(1966)，Helly(1911—1912)和 Bray(1918—1919)证明了 Helly-Bray 定理的形式(11.1.2).

显然，A. D. Alexandroff(1940)是第一个发展了依 L 收敛的一般定理，或者在相当一般的空上，有限测度的弱收敛性——事实上，这些空间比拓扑空间更普通，因为仅仅要求开集的可数并是开的(我不知道附加的一般性是否被真正的运用). Alexandroff(1940, p. 308)在没有更早的参考书的情况下给出了其定义.

portmanteau 定理(11.1.1)归功于 Alexandroff(1940, p. 312; 1943, 定理 2, p. 180). 他描述的情形如下:

变化好像是在一个空间里面游泳，奋力游向一个极限分布……变化在空间中可能进入一个闭集合 F，跑到了它的边界，如果不能在只有极限的情况下从里面跑出来，最后他们不得不从闭集合中跑出来，但不可能在极限的情况下发生，而只能在一个运动的特定阶段发生，因此在趋近于极限之前发生.

见 9.3 节的注释和 A. D. Alexandroff et al. (1973).

"portmanteau"这个词早期的意思是大的旅行箱，后来比喻为多个用途.

11.2 节 1864 年利普希茨定义了条件 $|f(x)-f(y)|\leqslant K|x-y|$(见 6.4 节的注释). Lipschitz(1876, p. 150)把$\mathbb{R}^k$ 上函数的这个条件与微分方程的系数和解联系起来. Liouville(1837, p. 19; 1838, p. 565)运用了解的逐次逼近的方法，在利普希茨条件下，Picard(1893)把这种方法推广到更一般的情况; Ince(1926, p. 76)给出了进一步的发展和参考文献. Lipschitz(1876, p. 157)也提到了条件 $|f(x)-f(y)|\leqslant K|x-y|^{\delta}$，其中 $0<\delta<1$. 对于任意的度量 d，d^{δ} 也是一个度量，因此这个条件也可以视为利普希茨条件. Hölder(1882)更好地运用了 $\delta<1$ 的这个条件，因此通常称为赫尔德条件.

命题 11.2.1 证明了 $BL(s, d)$是一个巴拿赫代数. 从这个方面它已经得到了广泛研究(Sherbert, 1964).

433 $BL(s, d)$及其性质的运用，至少在概率方面的应用，显然开始于 Fortet 和 Mourier(1953).

11.3 节 Prohorov(1956)定义了度量 ρ 且证明了它在完备可分度量空间中是可度量化依 L 收敛. Fortet 和 Mourier(1953)定义了度量 β，Ranga Rao(1962)证明了依 L 收敛与一致有界、等度连续函数族是一致的(推论 11.3.4)，因此推出了 β 是收敛的. Dudley(1966, 1968)证明了本节中的进一步结论.

11.4 节 Glivenko(1933)首先在 F 连续时证明了定理 11.4.2. Cantelli(1933)把它推广到一般的分布函数上(见《Jahrbuch》, 59: 1166 的摘要). Glivenko 生于 1897 年卒于 1940 年，他还致力于数学逻辑学、数学基础、实变函数和代数方面的研究(在 Istoriia, 1970, 其名字有 11 项索引). 关于坎泰利，见 8.3 节的注释(博雷尔-坎泰利引理). Varadarajan(1958)首先证明了定理 11.4.1.

11.5 节 在局部紧空间的测度理论中，对 μ 的定义域中的任意 A，测度 μ 称为是正规的，如果 $\mu(A)=\sup\{\mu(K): K\subset A, K$ 是紧的$\}$. (例如，Halmos, 1950, p. 224). 据我所知最早用到"胎紧"这个词的人是 Le Cam(1957)，是在讨论测度的集合时提到的，这里称为一致胎紧. Szpirajn(1937, 2.5(v))证明了如果(S, d)是一个可分度量空间，那么存在一个从 S 映上到另外一个度量空间 T 上的 1-1 博雷尔可测函数 f，其有博雷尔可测的逆，那么 S 是普遍可测的，当且仅当 T 是普遍可测的. 这个

事实改进了推论 11.5.2，其中 f 和 f^{-1} 是连续的．其英语的证明见 Shortt(1984)．Szpilrajn 的论文是用波兰语写的．我要感谢 Rae Shortt 承担此篇论文的审稿工作．后来 Szpilrajn 用 Mrczewski 这个名字发表文章．

定理 11.5.3 要归于 Le Cam(1957，定理 4，p.222)．在定理 11.5.4 中，由(Ⅱ)⇒(Ⅰ)时，假设 S 是一个波兰空间(对某个度量化是可分和完备的)不能被减弱．根据 D. Preiss(1973)，其含义对于所有非波兰可分度量化空间是不成立的，其是它们的完备化空间上的博雷尔集，例如，所有有理数的空间．回忆一个可分度量空间是波兰空间(对于有同样拓扑的其他度量是完备的)，当且仅当在它的完备化空间中，它是一个 G_δ(开集的可数交)(定理 2.5.4)．

11.6 节 P. Hall(1935)最早给出了配对定理(11.6.1)的明确阐述和证明．它很容易从 D. könig(1931，1950)的一个更早的定理得到，见 M. Hall(1958)．

Strassen(1965)运用哈恩-巴拿赫定理的形式(如 6.2 节中，凸集的分离)证明了对于 S 完备的主要定理(11.6.2)．这个定理的证明和在定理 11.6.2(Ⅱ)中用 $b>\beta$ 取代 β 的推广是 Dudley(1968)完成的．据我所知，用 β 取代 $b>\beta$ 和定理 11.63 在本书的第一版(1989)之前没有被完成．在线性规划方面，这个定理也有重要应用(Schay，1974)．

11.7 节 Persi Diaconis(个人信息)询问依 L 收敛的拓扑的一致相容性和怎样去刻画 ρ 和 β 的一致性．据我所知，定理 11.7.1 在本书第一版出版的时候(1989)已经完成．

定理 11.7.2 对于完备可分度量空间的完成要归于 Skorohod(1956)，关于可分度量空间要归于 Dudley(1968，定理 3)．Wichura(1970)和 Dudley(1985)通过限定 P 有可测的支撑把它推广到了非可分的度量空间上．Dudley(1966)处理了在拓扑空间上依 L 收敛的一致性．

434

对任意两个完备可分的度量空间 X 和 Y，以及不可数博雷尔集 $A\subset X$ 和 $B\subset Y$，如同在 13.1 节中所证的，存在一个从 A 映上到 B 的 1-1 博雷尔函数 f，其有博雷尔逆．运用这个稍微有些深奥的定理，首先对于 $[0,1]$ 上的某个法则 μ 得到 $P=\mu\circ f^{-1}$ 即可以证明定理 11.7.5，然后 $\mu=\lambda\circ X_F^{-1}$，其中 F 是 μ 的分布函数，X_F 由命题 9.1.2 给定的，所以有 $P=\lambda\circ(f\circ X_f)^{-1}$．定理 11.7.5 由 Halmos 和 Neumann(1942，定理 2)证得．Rokhlin(1949，§2，p.7)证明了 f 可以是 1-1 的．Dyson(1953)给出了在 $\mathbb{R}$ 上法则的例子，法则离得很远，但是如命题 11.7.6 中所证，其特征函数是一致近的．

11.8 节 Monge(1781)提出了下面的“运输问题”：在空间中给定两个区域 A 和 B，$V(A)=V(B)$，如何把 A 和 B 分解成任意小的部分，并且使 A 中的每一部分 A_i 与有相同容积的 B 中的每一部分 B_i 匹配，以致 $\sum_i V(A_i)d(A_i, B_i)$ 最小，其中 $d(A_i, B_i)$ 是从 A_i 到 B_i 的距离(当 A_i 和 B_i 趋近于 0 时，它是明确定义的)．最小值记作是 $cW(P, Q)$，其中 P 和 Q 分别是 A 和 B 上的一致概率分布，且 $c=V(A)$．这个问题与数学经济有关，这已在正文中提到．

L. V. Kantorovich 在 1939 年提出了线性规划——有联合线性约束的多元线性函数的最大值和最小值．例如，见 Kantorovich(1940)，他也用到了凸集的分离定理．关于 Kantorovich 在数学方面的工作见 Akilov et al(1962)、Vulikh et al(1972)和 Aganbegyan et al.(1987)．1975 年，Kantorovich 和 T. C. Koopmans 由于对“资源分配的最优化问题”的研究获得了诺贝尔经济奖——见 Scarf，1975. Kantorovich 生于 1912 年，卒于 1986 年(《New York Times》，1986)．

根据 Rachev(1984)的回忆，$\mathbb{R}$ 上关于法则的度量在最开始的时候是有限的，$W(P, Q):=\inf\{E|X-Y|: \mathcal{L}(X)=P, \mathcal{L}(Y)=Q\}$，有时称为 Gini 指标，其在 1914 年左右被 Gini 所研究．在

Salvemini(1943)对离散的分布作了证明之后，Dall'Aglio(1956，p. 42)由分布函数证明了 $W(P, Q)$ 的表示法．

Kantorovich 和 Rubinstein(Rubinshtein)(1958)首先证明了他们的定理(仅仅对紧的度量空间，但是在某些更一般的形式下)．据我所知，迄今为止还没有英文版出版．目前的证明是基于 J. Neveu(1974 年我从他那里学到了引理 11. 8. 5 和引理 11. 8. 6)．主题是关于无限维线性规划(Rubinshtein，1970，esp. p. 182—183，和 Vershik，1970)．Dudley 的不完善证明(1978，Lecture 20)被 de Acosta(1982，p. 368—372)完善．同时，Szulga(1978，1982)和 Fernique(1981)给出了另外一个证明．Rachev(1984)发现了这个问题并叙述了它．

这方面，我在与 A. de Acosta、J. Neveu、W. Philipp、S. T. Rachev、B. Simon 和 V. M. Zolotarev 的通信和讨论中受益匪浅．

11.9 节 Hoeffding(1948)证明了本节的主要结论(定理 11. 9. 3)．在较早的时候，Halmos(1946)已经证明了 U-统计量作为无偏估计的最优特性，同时 Von Mises(1947)得到了关于渐近分布的一些结果．

435

参考文献

de Acosta, Alejandro (1982). Invariance principles in probability for triangular arrays of B-valued random vectors and some applications. *Ann. Probability 10*: 346–373.

Aganbegyan, A. G., A. D. Aleksandrov, D. K. Faddeev, M. K. Gavurin, S. S. Kuteladze, V. L. Makarov, Yu. G. Reshetnyak, I. V. Romanovskii, G. Sh. Rubinshtein, and S. L. Sobolev (1987). Leonid Vital'evich Kantorovich. *Russian Math. Surveys 42*, no. 2: 225–232. Transl. from *Uspekhi Mat. Nauk 42*, no. 2: 177–182.

Akilov, G. P., B. Z. Vulikh, M. K. Gavurin, V. A. Zalgaller, I. P. Natanson, A. G. Pinsker, and D. K. Faddeev (1962). Leonid Vital'evich Kantorovich (In honor of his fiftieth birthday). *Russian Math. Surveys 17*, no. 4: 115–128. Transl. from *Uspekhi Mat. Nauk 17*, no. 4: 201–215.

Alexandroff, Alexander Danilovich [Aleksandrov, A. D.] (1940, 1941, 1943). Additive set-functions in abstract spaces. *Mat. Sbornik* (N.S.) *8*: 307–348; *9*: 563–628; *13*: 169–238 (in English).

Alexandroff [Aleksandrov], P. S., N. V. Efimov, V. A. Zalgaller, and A. V. Pogorelov (1973). Aleksandr Danilovich Aleksandrov (on his sixtieth birthday). *Russian Math. Surveys 28*, no. 2: 225–230. Transl. from *Uspekhi Mat. Nauk 28*, no. 2: 249–253.

Bray, H. E. (1918–1919). Elementary properties of the Stieltjes integral. *Ann. Math.* (Ser. 2) *20*: 177–186.

*Cantelli, Francesco Paolo (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari, 4*: 421–424.

Dall'Aglio, Giorgio (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuola Norm. Sup. Pisa* (Ser. 3) *10*: 35–74.

Dudley, R. M. (1966). Convergence of Baire measures. *Studia Math. 27*: 251–268.

——— (1968). Distances of probability measures and random variables. *Ann. Math. Statist. 39*: 1563–1572.

——— (1976). *Probabilities and Metrics*. Lecture Notes Ser. No. 45, Matematisk Institut, Aarhus Universitet.

——— (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Probability in Banach Spaces V* (Proc. Conf. Medford, 1984). Ed. A. Beck et al. *Lecture Notes in Math.* (Springer) *1153*: 141–178.

Dyson, Freeman J. (1953). Fourier transforms of distribution functions. *Canad. J. Math.* *5*: 554–558.

Fernique, Xavier (1981). Sur le théorème de Kantorovitch-Rubinstein dans les espaces polonais. *Séminaire de Probabilités XV. Lecture Notes in Math.* (Springer) *850*: 6–10.

Fortet, Robert, and Edith Mourier (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. Ecole Norm. Sup.* *70*: 266–285.

*Glivenko, Valery Ivanovich (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari* *4*: 92–99.

Hall, Marshall, Jr. (1958). A survey of combinatorial analysis. In *Some Aspects of Analysis and Probability*, pp. 35–104. Wiley, New York.

Hall, Philip (1935). On representatives of subsets. *J. London Math. Soc.* *10*: 26–30.

Halmos, Paul (1946). The theory of unbiased estimation. *Ann. Math. Statist.* *17*: 34–43.

——— (1950). *Measure Theory*. Van Nostrand, Princeton.

———, and von Neumann, J. (1942). Operator methods in classical mechanics, II. *Ann. Math.* *43*: 332–350.

*Helly, E. (1911–1912). Über lineare Funktionaloperatoren. *Sitzungsber. Nat. Kais. Akad. Wiss. (Wien) Math. Nat. Kl.* IIa *121*: 265–297.

Hildebrandt, T. H. (1966). Linear continuous functionals on the space (BV) with weak topologies. *Proc. Amer. Math. Soc.* *17*: 658–664.

Hoeffding, Wassily (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* *19*: 293–325.

*Hölder, Otto Ludwig (1882). Beiträge zur Potentialtheorie. Dissertation, Tübingen.

Ince, Edward Lindsay (1926). *Ordinary Differential Equations*. Repr. Dover, New York, 1956.

Istoriia otechestvennoi matematiki (in Russian) (1970). Vol. 4, bk. 2. Acad. Sci. USSR and Ukrain. SSR, Kiev.

Jahrbuch über die Fortschritte der Mathematik, 59 (1933).

Kantorovich, Leonid Vasilevich (1940). A new method of solving some classes of extremal problems. *Comptes Rendus (Doklady) Acad. Sci. USSR* *28*: 211–214.

———, and G. Sh. Rubinshtein (1958). On a space of completely additive functions. *Vestnik Leningrad Univ.* *13*, no. 7, *Ser. Mat. Astron. Phys.* *2*: 52–59 (in Russian).

*König, Dénes (1931). Graphs and matrices. (In Hungarian.) *Mat-Fiz. Lapok* *38*: 116–119.

——— (1950). *Theorie der endlichen und unendlichen Graphen*. Chelsea, New York.

Le Cam, Lucien (1957). Convergence in distribution of stochastic processes. *Univ. of Calif. Publs. in Statistics* *2*, no. 11: 207–236 (Univ. of Calif. Press, Berkeley).

Liouville, Joseph (1837). Second Mémoire: Sur le développement des fonctions ou parties de fonctions en séries dont les divers termes sont assujettis à satisfaire à une même équation différentielle du second ordre contenant un parametre variable. *J. de Math. Pures et Appliquées* (Ser. 1) *2*: 16–35.

——— (1838). Premier mémoire. Sur la théorie des équations différentielles linéaires et sur le développement des fonctions en séries. *J. de Math. Pures et Appliquées* (Ser. 1) *3*: 561–614.

Lipschitz, Rudolf Otto Sigismund (1876). Sur la possibilité d'intégrer complètement un système donné d'équations différentielles. *Bull. Sci. Math.* (Ser. 1) *10*: 149–159.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functionals. *Ann. Math. Statist.* *18*: 309–348.

*Monge, Gaspard (1781). Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie des Sciences de Paris, avec les Mémoires de mathématique et de physique pour la même année*, pp. 666–704.

New York Times [unsigned] (1986). Leonid V. Kantorovich dies; Won Nobel Economics Prize. April 11, 1986, p. D18.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

Picard, Émile (1893). Sur l'application des méthodes d'approximation successives a l'étude de certaines équations différentielles ordinaires. *J. de Math. Pures et Appliquées* (Ser. 4) *9*: 217–271.

Preiss, David (1973). Metric spaces in which Prohorov's theorem is not valid. *Z. Wahrscheinlichkeitsth. verw. Geb. 27*: 109–116.

Prohorov, Yurii Vasilevich (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appls. 1*: 157–214.

Rachev, Svetlozar T. (1984). The Monge-Kantorovich mass transference problem and its stochastic applications. *Theor. Probability Appls. 29*: 647–676.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist. 33*: 659–680.

Rokhlin, Vladimir Abramovich (1949). On the fundamental ideas of measure theory. *Mat. Sbornik 25*: 107–150. (In Russian); *Amer. Math. Soc. Translations 71* (1952): 1–54.

Rubinshtein, Gennadii Shlemovich (1970). Duality in mathematical programming and certain problems in convex analysis. *Russian Mathematical Surveys 25*: no. 5: 171–201.

*Salvemini, T. (1943). Sull calcolo degli indici di concordanza tra due caratteri quantitativi. In *Atti della VI Riunione della Soc. Ital. di Statistica*. Rome.

Scarf, Herbert E. (1975). The 1975 Nobel Prize in Economics: Resource allocation. *Science 190*: 649, 710, 712.

Schay, Geza, (1974). Nearest random variables with given distributions. *Ann. Probability* 2: 163–166.

Sherbert, Donald S. (1964). The structure of ideals and point derivations in Banach algebras of Lipschitz functions. *Trans. Amer. Math. Soc. 111*: 240–272.

Shortt, Rae Michael (1984). Universally measurable spaces: An invariance theorem and diverse characterizations. *Fund. Math. 121*: 169–176.

Skorohod, Anatolii Vladimirovich (1956). Limit theorems for stochastic processes. *Theor. Probability Appls. 1*: 261–290.

Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist. 36*: 423–439.

*Szpilrajn, E. (1937). On absolutely measurable sets and functions (in Polish). *C. R. Soc. Sci. Lett. Varsovie* CI. III *30:* 39–68.

Szulga, Alicja (1978). On the Wasserstein metric. *Trans. Eighth Prague Conf. Information Theory, Statistical Decision Functions Random Processes*, vol. B, pp. 267–273. Reidel, Dordrecht.

——— (1982). On minimal metrics in the space of random variables. *Theory Probability Appls.* 27: 424–430.

Varadarajan, Veeravalli S. (1958). On the convergence of sample probability distributions. *Sankhyā, 19*: 23–26.

Vershik, A. M. (1970). Some remarks on infinite-dimensional linear programming problems. *Russian Mathematical Surveys 25*, no. 5: 117–124.

Vulikh, B. Z., M. K. Gavurin, A. N. Kolmogorov, Yu. V. Linnik, V. L. Makarov, B. S. Mityagin, A. G. Pinsker, G. S. Rubinshtein, and D. K. Faddev (1972). Leonid Vitalevich Kantorovich (On his 60th birthday). *Russian Math. Surveys 27*, no. 3: 193–201. Transl. from *Uspekhi Mat. Nauk 27*, no. 3: 221–227.

Wichura, M. J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist. 41*: 284–291.

第12章　随机过程

迄今为止，随机过程的研究主要与鞅有关，虽然一般的定义已经给出：一个随机过程是两个变量 $t(t\in T)$ 和 $\omega(\omega\in\Omega)$ 的函数 X，其中$(\Omega,\ \mathcal{B},\ P)$是一个概率空间，且对每个 t，$X(t,\ \cdot)$在 Ω 上是可测的. 取 T 是正整数集合，任意随机变量序列都是一个随机过程. 在大多数过程的经典理论中，T 是实直线的子集. 但是到 20 世纪 50 年代，人们开始意识到存在着极不规则的随机过程，它们在表示和近似"噪音"方面有很大的用处，例如，以某种方式定义在直线上的随机过程，但此随机过程在任意点 t 处没有值. 相反，积分 $W(f)=\int W(t)f(t)\,\mathrm{d}t$ 是有定义的，如果 f 是光滑的和/或具有其他正则性质. 因此，对于过程的适当的指标集 T 可以是$\mathbb{R}$ 上函数的集合而不是$\mathbb{R}$ 的一个子集. 这样可以得到不仅随着时间变化而且在空间也有定义的随机函数，这样的过程也是很有用的，因此，T 也可以是空间的光滑函数的集合或者是有时间和空间变量的光滑函数的集合. 至少，对于任意没有结构的指标集 T，随机过程开始定理和基本存在定理都是成立的. 那么，最重要的维纳(Wiener)过程或布朗运动将被定义和研究. 如果 X_i 是独立同分布变量，其均值是 0、方差是 1，$S_n=X_1+\cdots+X_n$，$\limsup\limits_{n\to\infty} S_n/(2n\log\log n)^{1/2}=1$（重对数律）将在 12.5 节中通过运用到布朗运动来证明.

12.1　过程的存在性和布朗运动

给定一个随机过程 x：$\langle t,\ \omega\rangle\mapsto x(t,\ \omega)$，$t\in T$，$\omega\in\Omega$，$\Omega$ 上的函数 $x(t,\ \cdot)$通常记作 x_t，因此$x_t(\omega):=x(t,\ \omega)$. 令$\mathbb{R}^T$ 为 T 上所有实值函数的集合.

假设对每个有限集 $F\subset T$，在$\mathbb{R}^F$ 上可以给出法则 P_F，对于每一个 F，是否存在一个满足 $\mathcal{L}(\{x_t\}_{t\in F})=P_F$ 的过程 x？如果存在，P_F 将必须是相容的：当 $F\subset G$ 时，f_{GF} 是$\mathbb{R}^G$ 映上到$\mathbb{R}^F$ 的自然投影，$P_G\circ f_{GF}^{-1}=P_F$. 这个条件将被证明是充分的，取 $\Omega=\mathbb{R}^T$，有 $x(t,\ f):=f(t)$为随机过程.　439

本章在定理 12.1.2 及其证明之后仅需要考虑实值函数的情况. 因此，首先看到的是仅考虑$\mathbb{R}$ 而不考虑更一般的可测空间.

一个可测空间$(X,\ \mathcal{S})$，这里的$\mathcal{S}$是 X 的子集的 σ-代数，$\mathcal{S}$称为标准的当且仅当存在 X 的拓扑，对于这个拓扑$\mathcal{S}$是紧度量空间，且$\mathcal{S}$是博雷尔集的 σ-代数. 如果$(X,\ \mathcal{S})$是标准的，$(Y,\ \mathcal{U})$是另外一个可测空间，f 是从 X 映上到 Y 的具有可测逆的 1-1 可测函数，那么显然$(Y,\ \mathcal{U})$也是标准的. 例如，若 σ-代数$\mathcal{S}$的所有子集都是标准的，任意有这样的 σ-代数$\mathcal{S}$的可数集 X 要么是有限的，要么在$\mathbb{R}$ 上等价于一个可数空间到紧度量空间$\{0\}\cup\{1/n:\ n=1,\ 2,\ \cdots\}$的函数. 定理 13.1.1 将证明对于每一个可分度量空间 Y，若其是其完备化空间的博雷尔子集，那么它就是标准的(具有博雷尔 σ-代数). 因此，$\mathbb{R}^k$ 上任意具有博雷尔 σ-代数的博雷尔集是标准的. 现在将证明一个更简单的情况.

12.1.1 命题　$(\mathbb{R},\ \mathcal{B})$是标准的，其中$\mathcal{B}$是$\mathbb{R}$ 上博雷尔集的 σ-代数.

证明　记$\mathbb{R}$ 是$\{0\}$和所有半开区间$(n,\ n+1]$和$[-n-1,\ -n)$ $(n\in\mathbb{N})$的不相交并，记$[0,\ 1]$是$\{0\}$和半开区间$(1/(m+1),\ 1/m]$ $(m=1,\ 2,\ \cdots)$的不相交并. 因此，存在一个从$[0,\ 1]$映上到$\mathbb{R}$ 的 1-1 的博雷尔可测函数.　□

可测空间(X, $\mathcal{S}$)称为是普遍可测的(u. m.)，当且仅当$\mathcal{S}$对于 X 上的度量 d 是博雷尔 σ-代数，对于$\mathcal{S}$，如 11.5 节中所定义的那样，(X, d)是可分度量空间，且其是它的完备化空间的普遍可测子集(对于完备化空间的博雷尔 σ-代数上的每一个概率法则的完全测度是可测的). 如前所述，可分度量空间是普遍可测的，当且仅当它上的每一个法则都是胎紧的(定理 11.5.1). 显然，任意的标准可测空间都是普遍可测的.

令 T 是任意一个集合，假设对于每个 $t\in T$，$(S_t, \mathcal{B}_t)$是一个可测空间. 假设对于每一个有限集 $F\subset T$，在笛卡儿积 $S_F := \prod_{t\in F} S_t$ 上给定法则 P_F，它具有$\mathcal{B}_t$的积 σ-代数. 那么当 $F\subset G$ 时，我们有从 S_G 映上到 S_F 的自然投影 f_{GF}，且当 $F\subset G\subset T$ 和 G 是有限时，如果 $P_G\circ f_{GF}^{-1} = P_F$，那么给定的法则是相容的.

440

首先，假设对于任意的 t，$S_t = \mathbb{R}$ 有博雷尔 σ-代数. 尤其是当法则 P_F 是定义在一个有限维向量空间$\mathbb{R}^F$ 上时，P_F 称为是有限维分布(finite-dimensional distribution). 定义相容有限维分布函数的一种方法是在每一个 S_t 上指定一个法则 P_t，且令坐标是独立的，换句话说，对于任意的 $t\in F$，如同 P_t 的积测度那样定义每一个 P_F. 在那种情况下，对于可数的 T，下面的定理将由定理 8.2.2 得出.

12.1.2 定理(科尔莫戈罗夫) *对于任意的 T，普遍可测. 空间$(S_t, \mathcal{B}_t)_{t\in T}$和任意法则$\{P_F$: F 有限，$F\subset T\}$的相容族(在 S_F 上有 P_F)，那么对于所有有限的 $F\subset T$，在 S_T 上存在一个概率测度 P_T，满足 $P_T\circ f_{TF}^{-1} = P_F$.*

证明 同样可以假设每一个 S_t 都是可分度量空间且是它的完备化空间的普遍可测子集，有博雷尔 σ-代数. 令$\mathcal{A}$是 S_T 中的所有子集的集族，对于博雷尔集 $B\subset S_F$ 和有限集 F，有 $A = f_{TF}^{-1}(B)$. 回忆在两个或有限多个可分度量空间 S_T 的笛卡儿积中，积拓扑的博雷尔 σ-代数等于每一个 S_T 上的博雷尔 σ-代数的积 σ-代数(命题 4.1.7). 为证明$\mathcal{A}$是一个代数，首先由于 $A^c = f_{TF}^{-1}(B^c)$，所以$\mathcal{A}$包含了集合 A 的完备化空间. 对于并集，假设 $D = f_{TG}^{-1}(C)$，其中有限集 $G\subset T$，博雷尔集 $C\subset S_G$. 令 $H := F\cup G$，那么 $A\cup D = f_{TH}^{-1}(f_{HF}^{-1}(B)\cup f_{HG}^{-1}(C))\in\mathcal{A}$，因此$\mathcal{A}$是一个代数.

对于 S_F 中每个 $A = f_{TF}^{-1}(B)\in\mathcal{A}$，其中 B 是可测的，令 $P_T(A) := P_F(B)$. 为证明 P_T 是明确定义的，和上面一样假设 $A = D$. 那么

$$A = (f_{HG}\circ f_{TH})^{-1}(C) = f_{TH}^{-1}(f_{HG}^{-1}(C)) = f_{TH}^{-1}(f_{HF}^{-1}(B)).$$

由于 f_{TH}映上到 S_H，这蕴涵着 $f_{HG}^{-1}(C) = f_{HF}^{-1}(B)$，并且根据相容性可得，

$$P_F(B) = P_H(f_{HF}^{-1}(B)) = P_H(f_{HG}^{-1}(C)) = P_G(C).$$

为证明 P_T 有一个到由$\mathcal{A}$生成的 σ-代数的可数可加的扩张，只需证明 P_T 在$\mathcal{A}$上是可数可加的(定理 3.1.4).

首先容易给出$(S_t, \mathcal{B}_t)$是所有标准可测空间(例如，有博雷尔 σ-代数的$\mathbb{R}$)时的证明. 因此可以假设 S_t 是具有博雷尔 σ-代数的所有紧度量空间. 那么根据吉洪诺夫定理(2.2.8)可知，S_T 是紧的. 对于每一个 $A = f_{TF}^{-1}(B)\in\mathcal{A}$，根据定理 7.1.3 可知，

441

$$P_T(A) = P_F(B) = \sup\{P_F(K): K\text{ 是紧的}, K\subset B\}.$$

对于每一个紧集 $K\subset B$，$L := f_{TF}^{-1}(K)$是紧空间 S_T 的一个闭子集，因此可得其是紧的，且 $L\subset A$. 所以 $P_T(A) = \sup\{P_T(L)$: L 是紧的，$L\subset A\}$. 那么根据引理 10.2.4 知，P_T 在$\mathcal{A}$上是可数可加的，即对于标准空间 S_t，结论是成立的.

现在对于一般的普遍可测空间 S_t，如果 P_T 不是可数可加的，那么存在集合 $W_n \in \mathcal{A}$，$\varepsilon > 0$，满足 $W_n \downarrow \varnothing$，但是根据定理 3.1.1 知，对于所有的 n，$P_T(W_n) > \varepsilon$. 那么对于某个有限集 $F(n) \subset T$ 和博雷尔集合 $B_n \subset S_{F(n)}$，这里可以假设 $F(1) \subset F(2) \subset \cdots$，$W_n = f_{TF(n)}^{-1}(B_n)$. 如果 $m \leqslant n$，令 $f_{nm} := f_{n,m} := f_{F(n)F(m)}$. 那么

$$f_{TF(n)}^{-1}(B_n) \subset f_{TF(m)}^{-1}(B_m) = (f_{nm} \circ f_{TF(n)})^{-1}(B_m) = f_{TF(n)}^{-1}(f_{nm}^{-1}(B_m)).$$

因此 $B_n \subset f_{nm}^{-1}(B_m)$.

对每个 n，令 $P_n := P_{F(n)}$. 那么对每个 n，$P_T(W_n) = P_n(B_n)$. 根据定理 7.1.3 和定理 11.5.1，对于普遍可测空间，取 $K_n \subset B_n$，使得 $P_n(B_n \setminus K_n) < \varepsilon/3^n$ 成立. 令

$$C_n := \bigcap_{m=1}^{n} f_{nm}^{-1}(K_m).$$

那么对所有的 n，$C_n \subset B_n$，且由于 K_n 是紧的和 $\bigcap_{m<n} f_{nm}^{-1}(K_m)$ 是闭的，所以 C_n 是紧的. 现在有

$$B_n \setminus C_n = \bigcup_{m=1}^{n} B_n \setminus f_{nm}^{-1}(K_m) \subset \bigcup_{m=1}^{n} f_{nm}^{-1}(B_m \setminus K_m),$$

所以有

$$P_n(B_n \setminus C_n) \leqslant \sum_{m=1}^{n} \varepsilon/3^m < \varepsilon/2, \quad P_n(C_n) \geqslant \varepsilon - \varepsilon/2 = \varepsilon/2.$$

对于 $n \leqslant r$，有

$$f_{rn}^{-1}(C_n) = f_{rn}^{-1}\left(\bigcap_{m=1}^{n} f_{nm}^{-1}(K_m)\right) = \bigcap_{m=1}^{n} f_{rm}^{-1}(K_m) \supset C_r,$$

所以 $f_{rn}(C_r) \subset C_n$. 因此，如果 $j \leqslant n \leqslant r$，那么 $f_{rj}(C_r) = f_{nj}(f_{rn}(C_r)) \subset f_{nj}(C_n)$.

由于 $C_r \neq \varnothing$，所以 $f_{rj}(C_r) \neq \varnothing$. 令

$$D_n := \bigcap_{r \geqslant n} f_{rn}(C_r) = \bigcap_{r > n} f_{rn}(C_r).$$

那么 D_n 是一个非空紧集的递减序列集的交，它也是紧的和非空的. 选择 $x_1 \in D_1$. 由定理 2.2.12(b)可以推出

442

$$f_{n+1,n}(D_{n+1}) = f_{n+1,n}\left(\bigcap_{r>n} f_{r,n+1}(C_r)\right) = \bigcap_{r>n} f_{r,n}(C_r) = D_n.$$

所以 $\forall n$，由 $f_{n+1,n}(x_{n+1}) = x_n$ 可以递归地得到 $x_n \in D_n$. 那么对 $m > n$，$f_{mn}(x_m) = x_n$. 由于 $x_n \in D_n \subset S_{F(n)}$，令 x_n 对于 $u \in F(u)$ 有坐标 $x_n(u)$. 如果 $\forall n$，$u \in F(n)$，令 $x(u) := x_n(u)$，那么对 $u \in \bigcup_n F(n)$ 是明确定义的. 对于其他的 u，令 $x(u)$ 是 S_u 的任意点. 那么 $x \in S_T$ 且 $\forall n$，$f_{TF(n)}(x) = x_n$. 由于 $D_n \subset C_n \subset K_n \subset B_n$，故 $\forall n$，$x \in f_{TF(n)}^{-1}(B_n) = W_n$，矛盾. □

在定理 9.5.7 中及其后定义的正规法则 $N(m, C)$ 通常也称为高斯的. 一个随机过程 x 称为高斯的当且仅当对每个有限集 $F \subset T$，$\mathcal{L}(\{x_t\}_{t \in F})$ 是高斯的. 这些法则是由从 T 映射到 $\mathbb{R}$ 的均值函数 $m(t) := Ex_t$ 和从 $T \times T$ 映射到 $\mathbb{R}$ 的协方差函数 $C(s, t) := Ex_s x_t - m(s)m(t)$ 决定的. 当 $s = t$ 时函数值为非负数，当 $s \neq t$ 时函数值为 0 的函数就是协方差函数的一个可能的形式，这表明 x_t 和 x_s 是独立的. 下面是一个存在性定理.

12.1.3 定理 令 T 是任意集合，m 是任意从 T 映射到 $\mathbb{R}$ 的函数，C 是从 $T \times T$ 映射到 $\mathbb{R}$ 的任意函数，满足 $\forall s, t \in T$，$C(s, t) = C(t, s)$ 对任意的有限集 $F \subset T$，$\{C(s, t)\}_{s,t \in F}$ 是非负定矩阵. 那么存在

一个高斯过程$\{x_t\}_{t\in F}$，其均值函数是m，协方差函数是C.

证明 如果$\{y_t\}_{t\in T}$是均值为0、协方差为C的一个高斯过程，那么$x_t := y_t + m(t)$是均值为m、协方差为C的高斯过程. 因此可以取$m\equiv 0$.

对于任意的有限集$F\subset T$，令C_F是C到$F\times F$的限制. 考虑在$\mathbb{R}^F$上的正规法则$N(0, C_F)$(定理9.5.7). 如果$G\subset F$且t是$\mathbb{R}^G$上的线性函数，或者等价于具有通常内积的$\mathbb{R}^G$上的一个点，那么在$\mathbb{R}^F$上$t\circ f_{FG}$是在G上有t坐标在$F\setminus G$上是0的线性型. 那么有内积等式$(C_F(t\circ f_{FG}), t\circ f_{FG}) = (C_G(t), t)$. 由于每一个都有特征函数$\exp(-(C_G(t), t)/2)$，所以$N(0, C_F)\circ f_{FG}^{-1} = N(0, C_G)$(根据定理9.5.7；同时这也是命题9.5.12的一个特殊情况). 因此，法则$N(0, C_F)$对于有限集$F\subset T$形成一个相容族，并且应用科尔莫戈罗夫定理(12.1.2). □

443

定义 令$(H, (\cdot, \cdot))$是一个实内积空间，H上的**类正态过程**(isonormal process)是对于所有的$f, g\in H$，均值$EL(f)=0$和协方差$EL(f)L(g)=(f, g)$的高斯过程L.

12.1.4 定理 对任意实内积空间H，在H上存在一个类正态过程L. 任意这样的L都是线性的，即对任意的$f, g\in H$和$c\in\mathbb{R}$，有

$$L(cf+g) = cL(f) + L(g)\ \text{a. s.}$$

证明 其非负定的性质可以从内积的定义中得到，因此根据定理12.1.3可知，存在一个类正态过程L. 对于其线性性，如果扩展

$$E([L(cf+g) - cL(f) - L(g)]^2),$$

并且运用L的性质，那么可以得到

$$\begin{aligned}&\|cf+g\|^2 - 2c(cf+g,f) + c^2\|f\|^2 - 2(cf+g,g) + 2c(f,g) + \|g\|^2\\ =\ &\|f\|^2(c^2-2c^2+c^2) + (f,g)[2c-2c-2c+2c] + \|g\|^2[1-2+1] = 0,\end{aligned}$$

因此L是线性的. □

这里非形式地引入布朗运动. 悬浮在液体中的细小粒子持续地与液体中的分子碰撞并且改变方向. 在这个过程的数学理想化模型中，粒子在每一个时刻都在改变方向. 因此，如果$s<t<u$且X_t是粒子在t时刻的位置，那么增量$X_u - X_t$和$X_t - X_s$是独立的. 我们将只考虑一维过程，其中X_t是实值的，例如，X_t是一个三维运动的x坐标. 同时，过程对于时间是齐次的，因此对于$0\leqslant s<t$，$X_t - X_s$的法则仅依赖于差值$t-s$. 过程在0时刻是0，所以$X_0=0$，且对于所有的$t\geqslant 0$均值$EX_t=0$. 令$\sigma^2(t) := EX_t^2$. 现在假设所有的增量$X_{tj} - X_{t(j-1)}$ $(j=1, \cdots, n)$是独立的，所以对于所有的整数$m, n>0$和有理数$q\geqslant 0$，我们可以得到$\sigma^2(nt) = n\sigma^2(t)$，$\sigma^2(t/m) = \sigma^2(t)/m$和$\sigma^2(qt) = q\sigma^2(t)$. 同样，由于$X_t$是任意多个小独立同分布增量的和，根据适度条件(如同在林德伯格定理，9.6.1中的条件)，X_t有一个正态的分布. 在任意情况下，都可以假设$\mathcal{L}(X_t) = N(0, \sigma^2(t))$，且$X_t$的任意有限集有一个联合正态分布. 通过设$\sigma^2(1)=1$使这个过程正态化，因此对于所有的有理数$q\geqslant 0$，$\sigma^2(q)=q$. 对于任意的$s<t$，$\sigma^2(t) = \sigma^2(s) + \sigma^2(t-s) \geqslant \sigma^2(s)$，所以$\sigma^2(\cdot)$是非递减的，且对于所有的$t\geqslant 0$，$\sigma^2(t)=t$. 那么对于$0<s<t$，根据独立性有，$EX_s(X_t - X_s)=0$，因此，

444

$$EX_sX_t = EX_s(X_s + (X_t - X_s)) = EX_s^2 = s = \min(s,t).$$

均值$EX_t=0$和协方差EX_sX_t是由高斯过程的法则唯一决定的，由于根据定理9.5.1知，法则$N(0, C)$是由其特征函数$\exp((-Ct, t)/2)$(在定理9.5.6和定理9.5.7中给出)和协方差C唯一确定的. 在这种情况下，这个过程将称为布朗运动，稍后将给出更形式化的定义.

下一步是考察布朗运动作为一个t的函数的连续性. 在一个拓扑空间(对于布朗运动是半直线

$[0, \infty)$或是其子空间)上对于 t 定义的一个过程 x 称为是样本连续的(sample-continuous),当且仅当对于所有的 ω,函数 $t \mapsto x_t(\omega)$ 是连续的. 过程 x 称为是依概率连续的(continuous in probability),当且仅当 $t(n) \to t$ 时,$x_{t(n)}$ 依概率收敛到 x_t.

例如,令 Ω 是具有勒贝格测度的单位区间,令 $x(t, \omega) := 1_{\{t=\omega\}}$. 这个过程不是样本连续的;相反,$t \mapsto x(t, \omega)$ 在 $t=\omega$ 处总是不连续的. 另一方面,对于每个 t,$x(t, \omega)=0$a. s.. 因此,如果令 $y(t, \omega) \equiv 0$,那么 y 是样本连续的且有相同的有限维分布(如同在定理 12.1.2 中,P_F 对于 F 是有限的). 从而样本连续不依赖于有限维分布. 所以按照目前定义的布朗运动,不是每一个布朗运动都是样本连续的,但是下面将证明存在这样的布朗运动.

如果对有限集 F,存在一个具有给定的 P_F 的过程 x,使得 $t \mapsto x(t, \omega)$ 对于几乎所有的 ω 是连续的,那么在不改变法则 P_F 的条件下,可以假设它是样本连续的(对于所有的 ω),因为在这个函数可能不连续的零概率集上,我们可以通过改变 $x(t, \omega)$(如恒等于 0)来使它成为连续函数.

下面是布朗运动过程的形式的定义.

定义 *一个布朗运动或维纳过程是一个在 $T=[0, \infty)$ 上均值为 0、协方差为 $Ex_sx_t=\min(s, t)$ 的样本连续的高斯过程.*

与布朗运动有紧密联系的一个过程是*布朗桥*(Brownian bridge),它是定义在 $T=[0, 1]$ 上均值是 0、协方差是 $Ey_sy_t=s(1-t)(0 \leqslant s \leqslant t \leqslant 1)$ 的一个样本连续的高斯过程. 所以 $y_0=y_1=0$a. s.. 桥的两个端点是同水平的. 布朗桥是作为经验分布函数中心极限定理的一个极限出现的. 回忆(11.4 节)对于$\mathbb{R}$ 上的任意法则,其分布函数为 F,存在经验分布函数 F_n,其在$\mathbb{R}$ 上当 $n \to \infty$ 时一致收敛到 F(Glivenko-Cantelli 定理,11.4.2). 这是强大数定律的一种. 对于满足 $EX_1^2 < \infty$ 的普通的独立同分布随机变量 X_1,X_2,…我们可以得到 $S_n/n - EX_1 \to 0$a. s.. 为理解收敛率,下一步是中心极限定理(9.5.6),也就是 $n^{1/2}(S_n/n - EX_1)$ 依 L 收敛到一个正态分布. 同样,对于经验分布函数 F_n,自然要考虑 $n^{1/2}(F_n - F)$. 对于任意固定的 x,当 $n \to \infty$ 时,测度 $n^{1/2}(F_n - F)(x)$ 收敛到 $N(0, F(x)(1-F(x)))$. 事实上,对于 x 值的任意有限集,这些量的联合分布收敛到一个正态分布. 如同在引理 11.4.3 中那样,经验分布函数的研究可以归约为研究$[0, 1]$上的一致分布函数 G,记为 $G(x) := \max(0, \min(x, 1))$,其中 x 是任意的. 对于它们的经验分布函数 G_n,令 $\alpha_n(t) := n^{1/2}(G_n(t) - t)$,$0 \leqslant t \leqslant 1$,注意 $G(t)=t$. 由于 $G_n - G$ 是 n 个独立的 $G_1 - G$ 的平均,对于任意 s 和 t,$\alpha_n(s)$ 和 $\alpha_n(t)$ 的协方差对于所有的 n 是相同的. $n=1$ 时比较容易计算,对于 $0 \leqslant s \leqslant t \leqslant 1$,它等于 $s(1-t)$. 所以布朗桥过程是一个像所有的 α_n 一样具有相同协方差的高斯过程. 它是 α_n 依 L 收敛的极限,在本书里仅考虑有限多个 t 的值.

445

由于布朗运动和布朗桥的定义要求必须是样本连续的,这些过程的存在性不是很明显,但它们确实存在.

12.1.5 定理(维纳) *存在布朗运动 x 和布朗桥 y(它们都是样本连续的).*

证明 如果 x 是一个布朗运动,且 $y_t := x_t - tx_1$,$0 \leqslant t \leqslant 1$,这直接表明 y 是布朗桥. 所以只需对布朗运动证明这个定理. 下面将会用到正态分布的界.

12.1.6 引理 *对于任意的 $c>0$,*

(a)$N(0, 1)([c, \infty)) \leqslant (2\pi)^{-1/2}c^{-1}\exp(-c^2/2)$,且当 $c \to +\infty$ 时,式子两边是渐近的(它们的比收敛到 1).

(b)$N(0, 1)([c, \infty)) \leqslant \exp(-c^2/2)$.

注：在(a)中，对于大数 c 而言，其上界越精炼，不等式的两边越精确，比较简单的界(b)对很多理论过程都是不足的，当然对大数 c 也是如此.

446

证明 对于(a)，

$$N(0,1)([c,\infty)) = (2\pi)^{-1/2}\int_c^\infty \exp(-x^2/2)\,\mathrm{d}x$$
$$\leqslant (2\pi)^{-1/2}\int_c^\infty \frac{x}{c}\exp(-x^2/2)\,\mathrm{d}x$$
$$= (2\pi)^{-1/2}c^{-1}\exp(-c^2/2).$$

根据洛必达法则知，两边是渐近的. 现在对于 $c \geqslant (2\pi)^{-1/2}$，由(a)可以得到(b). 对于 $0<c<(2\pi)^{-1/2}$，

$$N(0,1)([c,\infty)) < N(0,1)([0,\infty)) = 1/2 < \exp(-1/(4\pi)) \leqslant \exp(-c^2/2),$$

即可证得(b). □

令 H 是希尔伯特空间 $L^2([0,\infty),\lambda)$，其中 λ 是勒贝格测度. 令 L 是 H 上的类正态. 对于 $0\leqslant t<\infty$，令 $x_t := L(1_{[0,t]})$，那么 $Ex_t=0$；对于 $0\leqslant s<\infty$，$Ex_sx_t=(1_{[0,s]},1_{[0,t]})=\min(s,t)$. 所以 x 有布朗运动的均值和协方差，它在任意有限集中对于 t 唯一确定 x_t 的联合分布函数. 由 L 的线性性(定理 12.1.4)可得，对于 $0\leqslant s\leqslant t$，$\mathcal{L}(x_t-x_s)=N(0,t-s)$. 下一个问题是寻找过程 z，使得对每个 t，$z_t=x_t$ a. s.，所以 z 的均值为 0，协方差为 $\min(s,t)$，并且 z 是样本连续的. 首先将要证明该结论对于 $0\leqslant t\leqslant 1$ 成立.

对每个 $n=1,2,\cdots$ 令 V_n 是所有等于 0 或，$x(j/2^n,\cdot)-x((j-1)/2^n,\cdot)$ $(j=1,\cdots,2^n)$ 的随机变量的集合. 那么对于除 $v=0$ 外的 $v\in V_n$，$\mathcal{L}(v)=N(0,2^{-n})$. 那么根据引理 12.1.6，

$$Pr\{\sup\{|v|:v\in V_n\}\geqslant n^{-2}\} \leqslant 2^nN(0,2^{-n})\{x:|x|\geqslant n^{-2}\}$$
$$\leqslant 2^{n+1}N(0,1)(2^{n/2}n^{-2},\infty) \leqslant P_n := 2^{n+1}\exp(-2^n/(2n^4)).$$

对于足够大的 n，$\exp(-2^n/2n^4)<4^{-n}$，所以 $\sum_n P_n<+\infty$. 根据博雷尔–坎泰利引理，对于几乎所有的 ω，存在一个 $n_0(\omega)$，使得对所有的 $n\geqslant n_0(\omega)$ 和 $v\in V_n$，有 $|v|\leqslant n^{-2}$.

对每个 $t\in[0,1]$，取二项展开 $t=\sum_{j\geqslant 1} t_j/2^j$，其中 $t_j=0$ 或 1. 对于二进制有理数 t，当 j 足够大时，选择 $t_j=1$ 的二项展开(除 $t=0$ 外). 令

$$t(n) := \sum_{j=1}^n t_j/2^j.$$

那么对每个 $n\geqslant 2$，$x_{t(n)}-x_{t(n-1)}\in V_n$. 由于 $\sum_n n^{-2}<\infty$，当 $n_0(\omega)$ 存在时，对于所有的 $t\in[0,1]$，极限 $z_t:=\lim_{n\to\infty}x_{t(n)}$ 是有定义的. 否则，集合 $z_t:=0$. 由于 $t\mapsto x_t$ 是依概率连续的，故对每个 t，$z_t=x_t$ a. s..(这里，概率为 0 的集合可能依赖于 t). 为证明 z 在 $[0,1]$ 上是样本连续的. 假设 $n\geqslant n_0(\omega)$ 且 $|s-t|\leqslant 2^{-n}$. 那么对于 $s(n)=j/2^n$ 和 $t(n)=k/2^n$，有 $|j-k|=0$ 或 1，所以当 $n\to\infty$ 时，

447

$$|z_s-z_t|\leqslant|z_{s(n)}-z_{t(n)}|+|z_s-z_{s(n)}|+|z_t-z_{t(n)}|\leqslant n^{-2}+2\sum_{m>n}m^{-2}\to 0.$$

因此 z 是样本连续的. 同样在 $[k,k+1]$ 上对于每个正整数 k 可以得到 $[o,\infty)$ 上的布朗运动是样本连续的. □

习题

1. 证明：在科尔莫戈罗夫存在性定理(12.1.2)中，普遍可测空间 S_t 可以由豪斯多夫空间来取代，在该豪斯多夫空间中所有的法则都是正则的(如 7.1 节中所定义的).
2. 在[0, 1]中，令 λ^* 是勒贝格测度 λ 的外测度，且令 A_n 是非可测集，如同在 3.4 节习题 2 中那样，对于所有的 n，满足 $A_n \downarrow \varnothing$ 和 $\lambda^*(A_n)=1$. 对每个 n，令 P_n 是由 $P_n(B\cap A_n):=\lambda^*(B\cap A_n)$ 定义的概率测度[提示：见定理 3.3.6]，其中 B 为博雷尔集. 对于每个 $x\in A_n$，令

$$f_n(x) := (x,x,\cdots,x) \in B_n := \prod_{j=1}^{n} A_j.$$

在 B_n 上令 $Q_n := P_n \circ f_n^{-1}$. 证明：$Q_n$ 在 $B_\infty = \prod_{j=1}^{\infty} A_j$ 上定义了一个有限维分布的相容族，但是在 B_∞ 上不存在有限维分布 Q_n 的概率测度.（因此，在习题 1 中的正则假设不能简单地去掉）.
3. 令 x_t 是布朗运动. 证明：对任意 $\varepsilon>0$ 和几乎所有的 ω，存在一个 $M(\omega)<\infty$，使得对于[0, 1]上的所有 s 和 t，$|x_s-x_t|\leqslant M(\omega)|s-t|^{0.5-\varepsilon}$. [提示：首先，在样本连续性(12.1.5)的证明中，用 $1/2^{n(1-\varepsilon)/2}$ 取代 $1/n^2$. 注意这些元素形成了关于 n 的几何级数，并且如果对每个 ω，$|s-t|$ 足够小，结论成立，那么对于[0, 1]中的所有 s 和 t 结论成立，其中 $M(\omega)$ 可能比较大.]
4. 注意到对于 $0\leqslant s\leqslant t\leqslant u\leqslant v$，$x_t-x_s$ 和 x_v-x_u 是独立的. 证明：对于任意的布朗运动 x_t，依概率地有 $\lim_{n\to\infty}\sum_{j=1}^{n}(X_{j/n}-x_{(j-1)/n})^2=1$. 求证在习题 3 中用 0. 5 + ε 代替 0. 5 − ε 时结论不成立. [提示：如果 $\mathcal{L}(X)=N(0,\sigma^2)$，那么 $EX^4=3\sigma^4$(命题 9.4.2).

448

5. 在命题 12.1.1 的证明中，一个从ℝ映上到[0, 1]的 1 − 1 可测函数 f 是由可数多个不相交的区间的并构成的，且令 f 是从ℝ上的每一个区间到[0, 1]中一个区间的单调函数. 证明：对于有限多个区间和半实线，命题 12.1.1 的结论不成立.
6. 运用导数证明：对于 $0<c<(2\pi)^{-1/2}$，引理 12.1.6(b)是成立的.
7. 对于 $a\geqslant 0$，$M(a):=\int_a^\infty \exp(-x^2/2)\,dx/\exp(-a^2/2)$. 这称为 Mills 率. 令 $f(a):=2((a^2+4)^{1/2}+a)^{-1}$ 和 $g(a):=2((a^2+a)^{1/2}+2)^{-1}$. 证明：对于所有的 $a\geqslant 0$，$f(a)\leqslant M(a)\leqslant g(a)$. [提示：证明 $f'\geqslant af-1$，$M'=aM-1$ 和 $g'\leqslant ag-1$，且这三个函数有上界 $1/a$. 考虑 $(M-f)'$ 并证明如果 $M-f$ 是负的且趋近于 $-\infty$，且该结论对于 $(g-M)'$ 同样成立.]
8. 令 F 是区间[3, 7]上一致分布的分布函数，求当 $n\to\infty$ 时，在$\mathbb{R}^2$上 $n^{1/2}(F_n(4)-F(4),\ F_n(6)-F(6))$ 依 L 收敛的极限.
9. 令 X_t 是任意随机过程，使得对于某个 $p\geqslant 1$ 和每个 $t\geqslant 0$，当 $S\to t$ 时，$E|X_s-X_t|^p\to 0$. 证明：$t\mapsto EX_t$ 是连续的.
10. 令 X_t 是一个高斯过程，其中 $0\leqslant t\leqslant 1$，其满足对于某个 $K<\infty$ 和 $0<p\leqslant 2$，对于[0, 1]上所有的 s 和 t，$E(X_t-X_s)^2\leqslant K|s-t|^p$. 证明：这个过程是样本连续的. 注意：假定 EX_t 不是 0 或常数.
11. 令 $T\subset\mathbb{R}$. 证明：在 T 上定义的依概率连续的过程仅依赖于有限维分布函数(与样本连续不同)；特别地，如果 x 和 y 是对于每一个有限集合 $F\subset T$ 使得 $\mathcal{L}(\{x_t\}_{t\in F})=\mathcal{L}(\{y_t\}_{t\in F})$ 成立的两个过程，那么 x 是依概率连续的，当且仅当 y 也是依概率连续的.
12. 回忆在ℕ上的泊松分布 P_λ，其参数 $\lambda\geqslant 0$，使得对于 $k=0,1,\cdots$，$P_\lambda(k)=e^{-\lambda}\lambda^k/k!$. 如果 X 有分布 P_λ，那么 $EX=\lambda$ 和 $\mathrm{var}(X)=\lambda$. 证明：存在一个过程 $p_t(t\geqslant 0)$，使得对于任意的 $0\leqslant s<t$，增量 p_t-p_s 有分布 p_{t-s}，且 p 有独立增量；即对于任意不相交区间 $(s-t]$ 的有限集合，p_t-p_s 是独立的. 那么证明过程 $v_t:=p_t-t$ 有布朗桥的均值和协方差，对于任意的 s 和 $t\geqslant 0$，$Ev_t\equiv 0$ 和 $Ev_sv_t=\min(s,t)$，但是不存在一个与 v 具有相同有限维联合分布的过程 z，使得 z 是样本连续的.

449　13. 明确定义一个 1－1 可测函数以完成命题 12.1.1 的证明.

12.2　布朗运动的强马尔可夫性质

在布朗运动过程 x_t 的研究中，需要研究其分布函数，例如，对于 $a>0$，$\sup_{0\le t\le a}|x_t|$ 或 $\sup_{0\le t\le a}x_t$. 尤其是要求事件 $\{\sup_{0\le t\le a}x_t\ge b\}$ 的概率，这里 $b>0$，注意到在这个事件中，至少存在一个时刻 τ，使得 $x_\tau\ge b(\tau\le a)$ 且根据连续性有 $x_\tau=b$. 那么考虑到增量 x_a-x_τ. 由于过程有均值为 0 的独立正态增量，看起来似乎 $P(x_a-x_\tau\ge 0\mid\tau\le a)=1/2$（这将在 12.3.1 中证明），那么

$$P(\sup_{0\le t\le a}x_t\ge b)=P(\tau\le a)=2P(x_a\ge b)=2N(0,a)([b,\infty])$$
$$=2N(0,1)(-\infty,-b/a^{1/2})=2\Phi(-b/a^{1/2})$$

这里的 Φ 是标准正态分布函数，即解得这个问题. 为完善这些细节和解决这类问题，布朗运动在马尔可夫时刻产生的一个"新起点"有着很好的作用，例如，产生的"新起点"τ. 在连续函数空间上，过程 $(x_{\tau+t}-x_\tau)_{t\ge 0}$ 与最初的布朗运动过程 $(x_t)_{t\ge 0}$ 有相同的法则. 这个性质称为强马尔可夫性质，这个性质作为本节的主要内容将在定理 12.2.7 中证明.

令 $x(t,\omega)(t\in T,\omega\in\Omega)$ 是概率空间 $(\Omega,\mathcal{S},P)$ 上的任意实值随机过程. 在 T 上所有实值函数的空间 $\mathbb{R}^T$ 上，定义估计函数或坐标函数 $e_t(f):=f(t)(\forall t\in T)$. 过程 x 定义一个从 Ω 映射到 $\mathbb{R}^T$ 上的函数 X: $\omega\mapsto x(\cdot,\omega)$，这个函数在 $\mathbb{R}^T$ 上对于最小 σ-代数 $\mathcal{B}^T$ 是可测的，并且对每个 $t\in T$，e_t 是可测的. 令 P_T 是 $\mathbb{R}^T$ 上的像测度 $P\circ X^{-1}$. 那么有一个过程 Y: $Y(t,f):=f(t)(t\in T,f\in\mathbb{R}^T)$，使得对于概率空间 $(\mathbb{R}^T,\mathcal{B}^T,P_T)$，任意的有限集 $F\subset T$，$\mathbb{R}^F$ 上 x 和 Y 的联合分布等于：

$$\mathcal{L}(\{Y(t,\cdot)\}_{t\in F})=\mathcal{L}(\{x(t\cdot)\}_{t\in F}).$$

这里 $\mathcal{B}^T$ 上的 P_T 也称为过程的法则 $\mathcal{L}(\{x_t\}_{t\in T})$. 对于 $\mathbb{R}^T$ 的任意包含每个函数 $x(\cdot,\omega)(\omega\in\Omega)$ 的子集，我们可以把 P_T 限制到 S 上.

令 $T:=[0,\infty)$，令 $C(T)$ 为 T 上连续实值函数的集合. 那么对于布朗运动过程 x，根据定理 12.1.5，集合 $S\subset\mathbb{R}^T$ 可以代替 $C(T)$.

定义 $C(T)$ 上的一个拓扑和度量是有用的. 那么对每个 $n=1,2,\cdots$ 定义 $C(T)$ 上的半范数 $\|\cdot\|_n$ 为

450
$$\|f\|_n:=\sup\{|f(x)|:0\le x\le n\}.$$

那么对每个 n，在 $C(T)$ 上给出伪测度 $d_n(f,g):=\|f-g\|_n$. 令 $h(t):=t/(1+t)$，$t\ge 0$，如同在命题 2.4.4 中那样，令 d 是由伪度量 d_n 定义的 $C(T)$ 上的度量，即 $d(f,g):=\sum_n h(d_n(f,g))/2^n$. 那么 $C(T)$ 中的序列 $\{f_m\}$ 对于 d 收敛，当且仅当它对于 d_n 是收敛的，即它在 T 的紧子集上是一致收敛的. 同样，序列 $\{g_j\}$ 对于 d 是一个柯西序列，当且仅当它对于每一个 d_n 和 n 是柯西序列. 那么它在 $[0,n]$ 上一致收敛到某个函数 $f_n\in C[0,n]$，满足当 $k\le n$ 时，在 $[0,k]$ 上 $f_k=f_n$. 因此对某个 $f\in C(T)$，对每个 n，在 $[0,n]$ 上 g_j 一致收敛到 f，所以 $d(g_j,f)\to 0$，即证得 $(C(T),d)$ 是完备的（这与定理 2.5.7 有关）. 由于每个 $C[0,n]$ 是可分的（推论 11.2.5），即可以推出 $C(T)$ 对于 d 是可分的. 另外，如果对于半范数序列定义的度量是完备的，则此向量空间称为弗雷歇(Fréchet)空间.

例如，如果 $f_n(t)=0$，$0\le t\le n$，且 $f_n(t\ge n)$ 是任意连续函数，那么不论 f_n 多大，在 $C(T)$ 对于 $t>n$ 都有 $f_n\to 0$.

令 $x(t,\omega)(t\in T,\omega\in\Omega)$ 是实值随机过程，这里的 $(T,\mathcal{U})$ 是一个可测空间，且 $(\Omega,\mathcal{C},P)$ 是

一个概率空间. 那么这个过程称为是可测的，当且仅当它是联合可测的，即对于积 σ-代数$\mathcal{U}\otimes\mathcal{C}$是可测的. 那么下面的事实是很有用的.

12.2.1 命题 如果(T, e)是一个可分度量空间，且x是样本连续的过程，那么x在T上对于博雷尔σ-代数是可测的.

证明 令$\{t_k\}_{1\le k<\infty}$在T中是稠密的. 对每个n和$k=1, 2, \cdots$令$U_{nk}:=\{t: e(t, t_k)<1/n\}$和$V_{nk}:=U_{nk}\setminus\bigcup_{j<k}U_{nj}$. 那么对每个$n$，$V_{nk}$是不相交的，且其并是$T$. 对每个$t\in V_{nk}$，$k=1, 2, \cdots$令$x_n(t, \omega):=x(t_k, \omega)$. 那么显然每个$x_n$都是联合可测的，并且对于所有的$t$和$\omega$，当$n\to\infty$时，$x_n(t, \omega)\to x(t, \omega)$. 那么$x$是可测的(根据定理4.2.2). □

把$C(T)$上对于d的博雷尔σ-代数$\mathcal{B}:=\mathcal{B}_{(d)}$和最小$\sigma$-代数$\mathcal{S}_T$联系起来是很有用的，其中对于$\mathcal{S}_T$，估计$e_t(e_t(f):=f(t))$对于每个$t\in T$是可测的.（在$\mathbb{R}^T$上对于积$\sigma$-代数$\mathcal{B}^T$，$\mathcal{S}_T=\{A\cap C(T): A\in\mathcal{B}^T\}$）.

12.2.2 命题 在$C(T)$上，$T:=[0, \infty)$，σ-代数$\mathcal{B}$和$\mathcal{S}_T$是相等的. $\mathcal{B}$上的任意法则P是由有限维分布$\{\mathcal{L}(\{f(t)\}_{t\in G}): G$有限，$G\subset T\}$的集合唯一确定的.

证明 每个e_t是连续的，那么e_t是博雷尔可测的，所以$\mathcal{S}_T\subset\mathcal{B}$. 反之，对于每一个$n=1, 2, \cdots[0, n]$上的上确界范数等于对$[0, n]$上所有有理数的可数集$Q_n$取上确界. 每一个闭球

451

$$\{f: \|f-g\|_n\le c\}, g\in C(T),\quad c\ge 0$$

是集合$\{f: |f(t)-g(t)|\le c\}(t\in Q_n)$的交. 那么这样的闭球在$\mathcal{S}_T$中，所以由$G_n(f):=d_n(f, g)$定义的函数$G_n$对于每一个$n$和$g\in C(T)$是$\mathcal{S}_T$可测的. 这可以推出$G(f):=d(f, g)$是可测的. 那么每一个开球$\{f: d(f, g)<r\}(g\in C(T), r>0)$在$\mathcal{S}_T$中. 由于$(C(T), d)$是可分的，根据命题2.1.4，对于$d$，每一个开集都在$\mathcal{S}_T$中，因此$\mathcal{B}=\mathcal{S}_T$.

现在正如在科尔莫戈罗夫定理(12.1.2)的证明中那样，由有限维联合分布确定的事件的集族是一个代数，它生成$\mathcal{S}_T$. 根据单调类定理(4.4.2)知，两个概率测度在同一个代数是一致的，同样这两个概率测度在基于该代数生成的σ-代数上也是一致的. □

对于任意的概率空间$(\Omega, \mathcal{A}, P)$，两个子σ-代数$\mathcal{B}\subset\mathcal{A}$和$\mathcal{C}\subset\mathcal{A}$是独立的，当且仅当$P(B\cap C)=P(B)P(C)$，其中$B\in\mathcal{B}$和$C\in\mathcal{C}$. 对于随机变量的一个集合$\mathcal{Y}$，令$\sigma(\mathcal{Y})$是最小的$\sigma$-代数，对于$\sigma(\mathcal{Y})$所有的$Y\in\mathcal{Y}$是可测的. 那么$\mathcal{Y}$称为与$\sigma$-代数$\mathcal{C}$是独立的，当且仅当$\sigma(\mathcal{Y})$与$\sigma$-代数$\mathcal{C}$是独立的. 如果$\mathcal{Z}$是另外一个随机变量的集合，那么$\mathcal{Y}$和$\mathcal{Z}$是独立的，当且仅当$\sigma(\mathcal{Y})$和$\sigma(\mathcal{Z})$是独立的. 通过等价类，这等价于$\mathcal{Y}$的任意有限子集与$\mathcal{Z}$的任意有限子集是独立的.

如果$\{x_t\}_{t\ge 0}$是一个布朗运动过程，且$0\le t_0<t_1<t_2<\cdots<t_n$，由于增量$x(t_j)-x(t_{j-1})(j=1, \cdots, n)$有联合正态分布且它们的协方差为0，所以它们是独立的. 特别地，它们的协方差矩阵的第j个对角项是t_j-t_{j-1}，其余项是0. 因此，根据定理9.5.14，对于每一个$t\ge 0$，所有$x_u(0\le u\le t)$的集合是与所有增量$x_{t+h}-x_t(h\ge 0)$的集合独立的.

在某个极限下保持独立性是很有用的.

12.2.3 命题 令$(\Omega, \mathcal{F}, P)$是一个概率空间，且$X_n(n=0, 1, \cdots)$是取值于一个可分度量空间S定义在Ω上的随机变量. 令$\mathcal{D}$是$\mathcal{F}$的一个子σ-代数，且对每个$n\ge 1$，令$\mathcal{A}_n$是一个σ-代数，对于$\mathcal{A}_n$，$X_1, \cdots, X_n$是可测的. 假设$\mathcal{A}_1\subset\mathcal{A}_2\subset\cdots$且每一个$\mathcal{A}_n$独立于$\mathcal{D}$. 令$X_n$依概率收敛到$X_0$，那么$X_0$也独立于$\mathcal{D}$.

452

证明　取一个子序列，根据定理 9.2.1，可以假设 $X_n \to X_0$ a. s.．令 $\mathcal{A}:=\bigcup_n \mathcal{A}_n$ 是一个代数．令 $\mathcal{J}=\{C\in\mathcal{F}: P(C\cap B)=P(C)P(B),\ \forall B\in\mathcal{D}\}$．那么 $\mathcal{J}$ 是包含 $\mathcal{A}$ 的单调类，所以根据定理 4.4.2，它包含由 $\mathcal{A}$ 生成的 σ-代数 $\mathcal{S}$．由于 X_0 几乎必然等于一个对于 $\mathcal{S}$ 可测的函数（根据定理 4.2.2），所以它独立于 $\mathcal{D}$． □

概率空间 $(\Omega,\mathcal{B},P)$，对于 $0\leqslant t\leqslant u$，满足 $\mathcal{B}_t\subset\mathcal{B}_u\subset\mathcal{B}$ 的 σ-代数的集合 $\{\mathcal{B}_t\}_{t\geqslant 0}$ 称为一个滤过 (filtration)．

定义　令 $(\Omega,\mathcal{B},P)$ 是一个概率空间，$\{x_t\}_{0\leqslant t<\infty}$ 是 Ω 上样本连续的布朗运动，且令 $\{\mathcal{B}_t\}_{0\leqslant t<\infty}$ 是一个滤过．那么 $\{x_t,\mathcal{B}_t\}_{t\geqslant 0}$ 是布朗运动当且仅当

(1) 对于 $0\leqslant t<\infty$，x_t 是 $\mathcal{B}_t$ 可测的，且

(2) 对于任意的 t，过程 $\{x_u-x_t\}_{u\geqslant t}$ 独立于 $\mathcal{B}_{t+}:=\bigcap_{v>t}\mathcal{B}_v$．

12.2.4 命题　如果 x_t 是一个样本连续的布朗运动，且 $\mathcal{F}_t$ 是最小的 σ-代数，对于 $\mathcal{F}_t$，$x_s(0\leqslant s\leqslant t)$ 是可测的，那么 $(x_t,\mathcal{F}_t)_{t\geqslant 0}$ 是布朗运动．

证明　如果在 (2) 中是 $\mathcal{B}_t$ 而不是 $\mathcal{B}_{t+}$，那么显然成立．对每个 $h>0$，$\{x_{u+h}-x_{t+h}\}_{u\geqslant t}$ 独立于 $\mathcal{B}_{t+h}\supset\mathcal{B}_{t+}$．那么通过某个序列令 $h\downarrow 0$，运用样本连续性，把命题 12.2.3 运用到 $S=C([t,\infty))$，结论成立． □

命题 12.2.4 中的 $\mathcal{F}_t$ 是给出布朗运动的最小可能的 σ-代数．例如，如果 X 是独立于 $\{x_t\}_{t\geqslant 0}$ 的随机变量，且 $\mathcal{C}_t$ 是包含 $\mathcal{F}_t$ 的最小 σ-代数，对于 $\mathcal{C}_t$，X 是可测的，那么 $(x_t,\mathcal{C}_t)_{t\geqslant 0}$ 也是布朗运动．

如果 $(x_t,\mathcal{B}_t)_{t\geqslant 0}$ 是布朗运动，那么对于 $(\mathcal{B}_t)_{t\geqslant 0}$，从 Ω 映射到 $[0,+\infty]$ 的函数 τ 称为停时 (stopping-time)，如果对于所有的 $t\geqslant 0$，$\{\tau\leqslant t\}\in\mathcal{B}_t$．或者函数 τ 称为马尔可夫时 (Markov time)，如果 $\forall t>0$，$\{\tau<t\}\in\mathcal{B}_t$．

首中时 (hitting time) $h_a:=\inf\{t: x_t=a\}\ (a\in\mathbb{R})$ 是一个停时，这将在 12.3.1 中被证明．

令 $(x_t,\mathcal{B}_t)_{0\leqslant t<\infty}$ 是任意布朗运动，如果 τ 对于 $(\mathcal{B}_t)_{t\geqslant 0}$ 是一个停时，令

453

$$\mathcal{B}_\tau:=\{A\in\mathcal{B}: A\cap\{\tau\leqslant t\}\in\mathcal{B}_t,\forall t\geqslant 0\}.$$

如果 τ 是一个马尔可夫时，令

$$\mathcal{B}_{\tau+}:=\{A\in\mathcal{B}: A\cap\{\tau<t\}\in\mathcal{B}_t,\quad\forall t>0\}.$$

那么显然 $\mathcal{B}_\tau$ 和 $\mathcal{B}_{\tau+}$ 是 σ-代数，注意到 $\Omega\in\mathcal{B}_t$ 当且仅当 τ 是一个停时，$\Omega\in\mathcal{B}_{\tau+}$ 当且仅当 τ 是一个马尔可夫时．

12.2.5 定理

(a) 如果 τ 是一个停时，那么它是 $\mathcal{B}_\tau$ 可测的．

(b) 如果 τ 是一个马尔可夫时，那么它是 $\mathcal{B}_{\tau+}$ 可测的．

(c) 如果 τ 是一个停时，那么它也是一个马尔可夫时，且 $\mathcal{B}_t\subset\mathcal{B}_{\tau+}$．

证明　(a) 对于任意的 $s\geqslant 0$ 和 $t\geqslant 0$，

$$\{\tau\leqslant s\}\cap\{\tau\leqslant t\}=\{\tau\leqslant\min(s,t)\}\in\mathcal{B}_{\min(s,t)}\subset\mathcal{B}_t.$$

所以对 $\forall s\geqslant 0$，$\{\tau\leqslant s\}\in\mathcal{B}_\tau$．那么由于对于 $0\leqslant s<\infty$，闭区间 $[0,s]$ 可以生成扩张到半直线 $[0,\infty]$ 的博雷尔 σ-代数，所以 (a) 成立．同理，由于 $\{\tau<s\}$，$\{\tau<t\}$，$0<s$，$t<\infty$，所以 (b) 成立．

对于 (c)，首先事件 $\{\tau<t\}$ 是事件 $T_n:=\{\tau\leqslant t-1/n\}\ (n=1,2,\cdots)$ 的并集，其中当 $t<1/n$ 时，

$T_n = \varnothing \in \mathcal{B}_t$，否则 $T_n \in \mathcal{B}_{t-1/n} \subset \mathcal{B}_t$. 对 n 取并，则 τ 是一个马尔可夫时. 令 $A \in \mathcal{B}_\tau$. 那么同理，对每个 n，$A \cap T_n \in \mathcal{B}_t$；由另外一个并得出 $A \in \mathcal{B}_{\tau+}$. □

滤过 $\{\mathcal{B}_t\}_{t\geq 0}$ 称为右连续的(right-continuous)，当且仅当对每一个 $t \geq 0$，

$$\mathcal{B}_t = \mathcal{B}_{t+} := \bigcap_{v>t} \mathcal{B}_v.$$

12.2.6 命题 对于一个右连续的滤过 $\{\mathcal{B}_t\}_{t\geq 0}$，$\tau$ 是马尔可夫时当且仅当它是一个停时，且 $\mathcal{B}_\tau = \mathcal{B}_{\tau+}$.

证明 根据定理12.2.5(c)可知，停时一定是一个马尔可夫时. 反之，令 τ 是一个马尔可夫时，对每个 $t \geq 0$，

$$\{\tau \leq t\} = \bigcap_{n=1}^{\infty} \{\tau < t + 1/n\}.$$

由于交是递减的，所以对于任意的 n_0，可以把它限制到 $n \geq n_0$ 上. 从而推出 $\{\tau \leq t\} \in \mathcal{B}_{t+} = \mathcal{B}_t$，所以 τ 是一个停时. 根据定理12.2.5(c)总是有 $\mathcal{B}_\tau \subset \mathcal{B}_{\tau+}$. 反之，令 $A \in \mathcal{B}_{\tau+}$，那么对于 $0 \leq t < u$，

$$A \cap \{\tau \leq t\} = A \cap \{\tau < u\} \cap \{\tau \leq t\} \in \mathcal{B}_u,$$

所以 $A \cap \{\tau \leq t\} \in \mathcal{B}_{t+} = \mathcal{B}_t$，于是 $A \in \mathcal{B}_\tau$. □ 454

由布朗运动 $(x_t, \mathcal{B}_t)_{t\geq 0}$ 的定义可以直接得出 $(x_t, \mathcal{B}_{t+})_{t\geq 0}$ 也是布朗运动. 显然滤过 $\{\mathcal{B}_{t+}\}_{t\geq 0}$ 是右连续的. 正如这里定义的，并不是所有的滤过都是右连续的. 所以不是所有的马尔可夫时都需要停时. 在某些点上，马尔可夫时可以交替地取具有右连续滤过的布朗运动，并且运用命题12.2.6.

如在下面的引理12.2.8和下一个例子中所示，σ-代数 $\mathcal{B}_{\tau+}$ 给出了 τ 的比邻信息.

例：按照下面的方法定义一个过程 z_t. 令 $m(\omega)$ 是 $[0, 1]$ 上的一致随机时间分布. 令 $s = s(\cdot)$ 是与 $m(\cdot)$ 独立的随机变量，且满足 $P(s=1) = P(s=-1) = 1/2$. 当 $s=-1$ 时，令 $z_t := m(\omega) - t$；当 $s=1$ 时，令 $z_t := |m(\omega) - t|$（见图12-2）. 令 $\mathcal{C}_t$ 是最小 σ-代数，对于 $\mathcal{C}_t$ 和 $0 \leq u \leq t$，z_u 是可测的. 那么 m 对于 σ-代数 $\mathcal{C}_t$ 是一个停时，并且对于 $s(\cdot)\mathcal{C}_{m+}$ 是可测的，当其对于 $\mathcal{C}_m$ 不可测时，即可证得对过程 z_t，两个 σ-代数的不同. S 的值不能由 $z_t(t \leq m(\omega))$ 得到，但是可证得它比邻 $m(\omega)$.

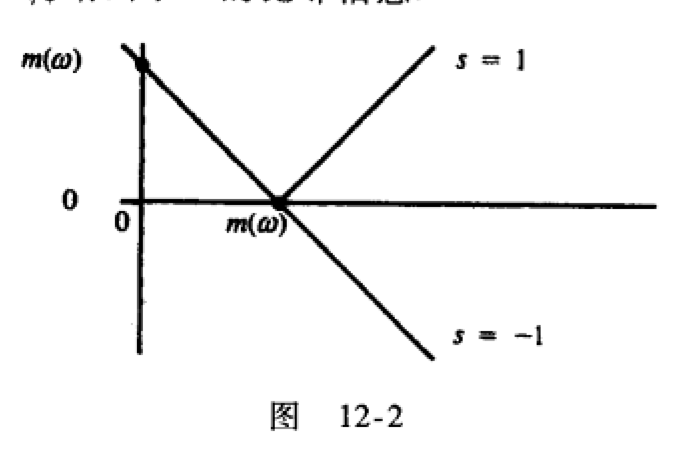


图 12-2

令 $(\Omega, \mathcal{A}, P)$ 是一个概率空间，且令 $v(t, \omega)(0 \leq t < \infty, \omega \in \Omega)$ 是一个样本连续的实值随机过程. 令 $V: \omega \mapsto v(\cdot\omega)$ 是从 Ω 映射到 $C([0, \infty))$ 的函数. 如命题12.2.2所示，在博雷尔 σ-代数 $\mathcal{B} = \mathcal{S}_{[0,\infty)}$ 上，令 $\mathcal{L}(v) := \mathcal{L}(V) := P \circ V^{-1}$.

如果 $A \in \mathcal{A}$ 且 $P(A) > 0$，对于 $E \in \mathcal{A}$，令 $P_{\restriction A}(E) := P(E \mid A) := P(E \cap A)/P(A)$. 那么 $P_{\restriction A}$ 在 $\mathcal{A}$ 上是一个概率测度. 令 $\mathcal{L}(v \mid A) := P_{\restriction A} \circ V^{-1}$. 如果 $\mathcal{D}$ 是 $\mathcal{A}$ 的一个子 σ-代数，那么 $v(\cdot,\cdot)$ 独立于 $\mathcal{D}$，当且仅当 $\{v(t, \cdot)\}_{t\geq 0}$ 独立于 $\mathcal{D}$，这就等价于 $\mathcal{L}(v \mid A) = \mathcal{L}(v)$，其中 $P(A) > 0$，$A \in \mathcal{D}$.

对于任意固定的 $t > 0$ 和布朗运动 x，显然 $\{x_{t+h} - x_t\}_{h\geq 0}$ 与 $x = \{x_h\}_{h\geq 0}$ 有相同的有限维法则（对于 h 限制到有限集），且在 $(C([0, \infty)), \mathcal{B})$ 上有相同的法则. 根据有滤过的布朗运动的定义知，对于每一个 i，在 $\{\tau = t(i)\} \in \mathcal{B}_{t(i)+}$ 上 $\{x_{t(i)+h} - x_{t(i)}\}_{h\geq 0}$ 的条件法则和在布朗运动上的法则是相同的，所以用一个仅有有限多个不同的(非负实数)值 $t_i := t(i)$ 的马尔可夫时 τ 代替 t，结论同样成立. 在这种意义下，在这些马尔可夫时上，布朗运动有一个全新的开始. 把这个事实推广到一般的马尔可夫时是很有用的. 455

12.2.7 定理(布朗运动的强马尔可夫性) 令$(x_t, \mathcal{B}_t)_{t\geqslant 0}$是任意布朗运动，且$\tau$对于$(\mathcal{B}_t)_{t\geqslant 0}$是一个马尔可夫时，满足$P(\tau<\infty)>0$. 那么在集合$\{\tau<\infty\}$上，$\omega\mapsto x_\tau:=x(\tau(\omega),\omega)$是$\mathcal{B}_{\tau+}$可测的. 对于$h\geqslant 0$，如果$\tau(\omega)<\infty$，令

$$W(h,\omega):=W_h(\omega):=x(\tau(\omega)+h,\omega)-x(\tau(\omega),\omega).$$

否则，没有定义. 那么对于任意满足$P(C)>0$的$C\in\mathcal{B}_{\tau+}$，在C上$\tau<\infty$，条件法则$\mathcal{L}(W\mid C)=\mathcal{L}(x)$，其中$x=\{x_t\}_{t\geqslant 0}$是布朗运动. 换句话说，对于$F:=\{\tau<\infty\}\in\mathcal{B}_{\tau+}$，在概率空间$(F, P_{\restriction F})$上的过程$W$是一个布朗运动且独立于$\sigma$-代数$\mathcal{B}^F_{\tau+}:=\{B\cap F: B\in\mathcal{B}_{\tau+}\}$.

证明 首先，下面的这些引理是有用的.

12.2.8 引理 如果τ是一个常数c，那么$\mathcal{B}_{\tau+}=\mathcal{B}_{c+}:=\bigcap_{t>c}\mathcal{B}_t$.

证明 首先注意到常量是一个停时，因此也是一个马尔可夫时. 如果$A\in\mathcal{B}_{\tau+}$，那么由$\mathcal{B}_{\tau+}$的定义可以得出$A\in\mathcal{B}_t(t>c)$，所以$A\in\mathcal{B}_{c+}$. 反之，如果$A\in\mathcal{B}_{c+}$，那么当$t>c$时，$A_t:=A\cap\{\tau<t\}=A\in\mathcal{B}_t$；否则$A_t=\varnothing\in\mathcal{B}_t$. 所以$A\in\mathcal{B}_{\tau+}$. □

如果$u\leqslant v$，且我们知道一个事件在时刻u是否发生，就会知道在时刻v该事件是否会发生，那么下一个事实是显然的.

12.2.9 引理 如果u和v是马尔可夫时，且$u\leqslant v$，那么$\mathcal{B}_{u+}\subset\mathcal{B}_{v+}$.

证明 如果$A\in\mathcal{B}_{u+}$，那么对于任意的$t>0$，

$$A\cap\{v<t\}=(A\cap\{u<t\})\cap\{v<t\}\in\mathcal{B}_t. \qquad \square$$

为继续强马尔可夫性的证明(定理 12.2.7)，令$\tau_n(\omega):=\tau(n)(\omega):=k/2^n$当且仅当$(k-1)/2^n\leqslant\tau(\omega)<k/2^n$，其中$k=1, 2, \cdots, n=1, 2, \cdots$，或者$\tau_n=+\infty$当且仅当$\tau=+\infty$. 那么由于对$k/2^n\leqslant t<(k+1)/2^n$，

[456]

$$\{\tau_n\leqslant t\}=\{\tau<k/2^n\}\in\mathcal{B}_t,$$

那么τ_n是停时，而对$t<1/2^n$，$\{\tau_n\leqslant t\}=\varnothing\in\mathcal{B}_t$.

如果$\tau<\infty$，那么对$\forall n$，有$\tau_n\downarrow\tau$且$\tau_n>\tau$. 要证明x_τ是$\mathcal{B}_{\tau+}$可测的，只需证明(根据定理 4.1.6)对每个实数y和$t>0$，$\{x_\tau>y\}\cap\{\tau<t\}\in\mathcal{B}_t$. 这个事件等价于：对于某些$m=1, 2, \cdots$和所有的$n\geqslant m$，$\tau(n)<t$和$x_{\tau(n)}\geqslant y+(1/m)$. 现在$\tau(n)$仅有可数多个可能的值$t(n, k)$，且每一个事件$\{\tau(n)=t(n,k)<t\}\cap\{x_{t(n,k)}\geqslant y+(1/m)\}$都在$\mathcal{B}_t$中，其中对于$t(n,k)<t$，根据定理 12.2.5(a)，$\{\tau(n)=t(n,k)\}\in\mathcal{B}_{\tau(n)}$，所以$x_\tau$是$\mathcal{B}_{\tau+}$可测的.

如果$\tau(\omega)<\infty$，令$W_n(h,\omega):=x(\tau_n(\omega)+h,\omega)-x(\tau_n,\omega)$；否则，没有定义. 那么根据样本连续性知，对于$\tau<\infty$，$h\geqslant 0$和$\omega$，有$W_n(h,\omega)\to W(h,\omega)$. 由于每一个函数$x(\cdot,\omega)$在紧集上是一致连续的，且$W_n$一致收敛到$W$. 那么当$n\to\infty$时，对于任意的$\omega\in F$，$d(W_n(\cdot,\omega), W(\cdot,\omega))\to 0$. 根据命题 12.2.1 知，在$\{\tau<\infty\}$上$W$是可测的.

如果τ是一个常数，根据引理 12.2.8 和有滤过的布朗运动的定义可以得出结论成立.

下面令$c:=c(n,k):=k/2^n$，对于固定的k，n和c，如果$A\in\mathcal{B}_{c+}$且$P(A\cap\{\tau_n=c\})>0$，那么根据“由于$\{\tau_n=c\}\cap A\in\mathcal{B}_{c+}$，故$\tau$是常数”的事实有$\mathcal{L}(W_n\mid A\cap\{\tau_n=c\})=\mathcal{L}(x)$. 如果$C\in\mathcal{B}^F_{\tau(n)+}$和$P(C)>0$，令$C_k:=C\cap\{\tau_n=c(n,k)\}$. 那么$C_k\in\mathcal{B}_{c(n,k)+}$和$C=\bigcup_k C_k$. 由于$W_n$独立于每一个$C_k$，且它们有同样的条件法则$\mathcal{L}(x)$，对其并$C$也同样成立. 所以根据引理 12.2.9 知，$W_n$独立于任意的$C\in\mathcal{B}^F_{\tau+}$. 对于任意使得$P(C)>0$成立的$C$和$C[0,\infty)$上的任意有界连续实值函数$f$，

$$E(f(W_n) \mid C) = \int_C f(W_n)dP/P(C) = Ef(x).$$

由于对于 d，W_n 收敛到 W，$Ef(W \mid C) = Ef(x)$，所以根据引理 9.3.2 知，在 $C[0,\infty)$上有 $\mathcal{L}(W \mid C) = \mathcal{L}(x)$. 这个法则不依赖于 C，所以在$(F, P_{\upharpoonright F})$上 W 是独立于$\mathcal{B}_{\tau+}^F$的. 即定理 12.2.7 得证. □

习题

在习题 1 ~ 3 和习题 5 ~ 6 中，如同在命题 12.2.4 中一样，其中的布朗运动具有最小 σ-代数$\mathcal{F}_t$. 457

1. 考虑一个布朗运动，在它的概率空间上定义一个随机变量 $\tau \geq 0$，这个变量 τ 不是一个马尔可夫时，且对某个 $h>0$，$x_{\tau+h} - x_\tau$ 与 x_h 没有相同的法则. [提示：令 $\tau := \inf\{t: x_{t+1} - x_t \geq 0\}$，$h = 1$.]
2. 对于一个布朗运动$\{x_t\}_{t\geq 0}$，令 $s(\omega) := \inf\{t: x_t > 1\}$ 和 $\tau(\omega) := \inf\{u > s(\omega): x_u < 0\}$，证明：$\tau$ 是马尔可夫时.
3. 对于布朗运动，给出一个马尔可夫时的例子，使得对于 $n \leq \tau(\omega) < n+1$，$n = 0, 1, \cdots$ 如果 $s(\omega) = n$，那么 s 不是一个马尔可夫时.
4. 如果对于一个滤过$\{\mathcal{B}_t\}_{t\geq 0}$，$\rho$ 和 τ 是两个马尔可夫时，证明：$\rho + \tau$ 对于$\{\mathcal{B}_t\}_{t\geq 0}$也是马尔可夫时.
5. 对于布朗运动，给出两个停时 ρ 和 τ，满足 $\tau < \rho < \infty$ a. s.，使得 $x_\rho - x_\tau$ 不独立于$\mathcal{B}_{\tau+}$. [提示：令 $\tau \equiv 1$，$\rho := \inf\{t > 1: x_t = 0\}$.]
6. 证明：对于 $b > 0$ 和 $c > 0$，$c + |x_b|$ 对于布朗运动是一个马尔可夫时当且仅当 $c \geq b$.
7. 令 $X_1, X_2, \cdots$是独立同分布实值随机变量，且 $S_n := X_1 + \cdots + X_n$. 令$\mathcal{B}_n$是最小的 σ-代数，对于$\mathcal{B}_n$，$X_1, \cdots, X_n$ 是可测的. 对于$\{\mathcal{B}_n\}_{n\geq 0}$，令 $k := k(\cdot)$是正整数停时，这就意味着对于所有的 $n = 1, 2, \cdots$，$\{\omega: k(\omega) \leq n\} \in \mathcal{B}_n$. 证明：$\{S_{n+k} - S_k\}_{n\geq 1}$独立于$\mathcal{B}_k := \{A: A \cap \{k \leq n\} \in \mathcal{B}_n, \forall n = 1, 2, \cdots\}$，并且它们和序列$\{S_n\}_{n\geq 1}$有同样的分布.
8. 证明：在 σ-代数 $C[0,\infty)$上，对于在紧区间上可度量化一致收敛的度量 d，布朗运动的法则是唯一的. (这个事实已经在正文中假设过). [提示：运用命题 12.2.2.]
9. (a)对于一个滤过$\{\mathcal{B}_t\}_{t\geq 0}$，如果 $\tau_n (n = 1, 2, \cdots)$是停时，且当 $n\to\infty$ 时，$\tau_n \uparrow \tau$，证明：对于$\{\mathcal{B}_t\}_{t\geq 0}$，$\tau$ 也是一个停时.

   (b)在(a)中把停时换成马尔可夫时结论同样成立. [提示：证明对 $t > 0$，$\{\tau < t\} = \bigcup_{q\in\mathbb{Q}, q<t} \bigcap_n \{\tau_n < q\}$.]
10. 证明：对于在图 12-2 的例子中的过程 z_t 和停时 m，$z_{m+h} - z_m$ 独立于$\mathcal{B}_m$，但它不独立于$\mathcal{B}_{m+}$(与布朗运动的强马尔可夫性相反).
11. 证明：对于一个滤过$\{\mathcal{B}_t\}_{t\geq 0}$，如果 τ 是一个马尔可夫时，那么对于$\{\mathcal{B}_{t+}\}_{t\geq 0}$，$\tau$ 是一个停时.
12. 令 $s = \pm 1$，其等于 1 和 -1 的概率都是 1/2，且 $X_t(\omega) := s(\omega)1_{(1,2]}(t)$. 令$\{\mathcal{B}_t\}$是最小 σ-代数，对于$\{\mathcal{B}_t\}$，$X_u (0 \leq u \leq t)$是可测的. 令 $\tau := 2 - s$. 证明：对于$\{\mathcal{B}_t\}_{t\geq 0}$，$\tau$ 是一个马尔可夫时但不是一个停时. 458

12.3 反射原理、布朗桥和上确界定律

令 X_n 是独立的，且它等于 1 或 -1 的概率都为 1/2(“掷硬币”随机变量)，且 $S_n := X_1 + \cdots + X_n$. S_n 称为一个简单随机游动. 注意到 n 和 S_n 有相同的奇偶性(同是奇数或同是偶数). 令 $0 < k < n$，这里的 n 和 k 有不同的奇偶性. 令 A_{kn}是偶数，对于 $j \leq n$，$S_j \geq k$. 在 A_{kn}上，对于这些 j，令 $m(\omega)$是最小的，那么由于 $n - m(\omega)$ 是奇数，所以 $S_n - S_{m(\omega)}$ 不可能是 0. 对每个 $j < n$，条件概率 $P(S_n - S_{m(\omega)} > 0 \mid m(\omega) = j) = 1/2$. 由于事件 $m(\omega) = j (j = 1, \cdots, n)$与 A_{kn}的并是不相交的，故有

$P(S_n - S_{m(\omega)} > 0 \mid A_{kn}) = 1/2$. 因此 $P(S_n > k) = P(S_n \geqslant k) = P(A_{kn})/2$，所以 $P(A_{kn}) = 2P(S_n > k)$. 这个事实就是“Désiré André 反射原理”. 这个原理的思想就是对于任意可能的路径的集合，在时刻 $m(\omega)$ 开始，直线 $y = k$ 中的反射保持概率.（见图 12-3A）.

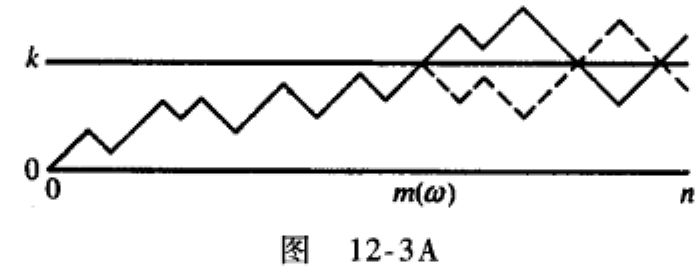


图 12-3A

在本节中，在某些有限区间上布朗运动的上确界的分布和布朗桥及它们的绝对值可以根据反射方法计算得出. 首先，下面是把 André 反射原理扩展到布朗运动上.

12.3.1 反射原理 令 $(x_t, \mathcal{B}_t)_{t\geqslant 0}$ 是布朗运动，且 $b, c > 0$. 那么

$$P(\sup\{x_t : t \leqslant b\} \geqslant c) = 2P(x_b \geqslant c) = 2N(0,b)([c,\infty)).$$

证明 令 $\tau := \tau(\omega) := \inf\{t: x_t = c\}$. 那么首中时 τ 是一个停时（可能是无限的），因为根据样本连续性，对于 $t < \infty$，$\{\tau \leqslant t\} = \bigcap_{q<c} \bigcup_{r<t} \{x_r > q\}$，其中 r 和 q 限制为有理数. $\sup\{x_t: t \leqslant b\} \geqslant c$ 当且仅当 $\tau \leqslant b$. 令 $B := \{\tau \leqslant b\}$. 那么根据命题 12.2.5 知，$P(B) > 0$，$B \in \mathcal{B}_\tau \subset \mathcal{B}_{\tau+}$，且在 B 上 $\tau < \infty$，所以可以运用强马尔可夫性（定理 12.2.7）. 令 W_h 是定义在事件 $\{\tau < \infty\}$ 上的布朗过程 $x_{\tau+h} - x_\tau$. 那么 $\{x_b > c\} \subset B$ 且

459

$$P(x_b \geqslant c \mid B) = P(x_b > c \mid B) = P(W_{b-\tau} > 0 \mid B).$$

在 B 上令 $s := b - \tau$. 那么有 $P(s = 0) \leqslant P(x_b = c) = 0$，所以在 B 上 $s > 0$a. s.. 由于 W 是样本连续的，根据命题 12.2.1 知，$\omega \mapsto W_{s(\omega)}(\omega)$ 是可测的.

令 $\mathcal{A}$ 是 B 的子集的最小 σ-代数，对于 $\mathcal{A}$，$W_t(t \geqslant 0)$ 限制于 $\omega \in \mathcal{B}$ 是可测的. 令 $\mathcal{B}'_{\tau+} := \{C \cap B: C \in \mathcal{B}_{\tau+}\}$. 对于任意事件 $D \subset B$，令 $P_B(D) := P(D)/P(B)$. 令 B' 和 B'' 是分别具有不同 σ-代数 $(B', \mathcal{B}'_{\tau+})$ 和 $(B'', \mathcal{A})$ 的集合 B. 令 u 和 v 分别是从 B 映射到 B' 和 B'' 的恒等函数. 那么对于给定的 σ-代数，根据强马尔可夫性知，对于 P_B，u 和 v 是独立的，所以在 $B' \times B''$ 上，(u, v) 的法则是 $\mu \times \rho$，它们法则的积. 记 $s = s(u)$ 和 $W_t(\omega) = W_t(v)$，其中 $t \geqslant 0$. 那么根据 Tonelli-Fubini 定理，

$$P(W_{s(\omega)} > 0 \mid B) = \int \rho(\{v: W_{s(u)}(v) > 0\})\,\mathrm{d}\mu(u) = \int \frac{1}{2}\mathrm{d}\mu(u) = \frac{1}{2}.$$

那么 $P(x_b \geqslant c \mid B) = 1/2$，$P(x_b \geqslant c) = P(B)/2$. □

布朗桥过程 $y_t (0 \leqslant t \leqslant 1)$ 是高斯过程，其均值是 0，协方差是 $Ey_s y_t = s(1-t)$，其中 $0 \leqslant s \leqslant t \leqslant 1$. 那么这个过程的上确界分布和它的绝对值的分布将在命题 12.3.3 和命题 12.3.4 中给出. 这些分布可以运用到下面的统计中.

正如在定理 12.1.5 前所注的，y_t 由过程 $n^{1/2}(G_n(t) - G(t))$ 的极限生成，这里的 G_n 是 $0 \leqslant t \leqslant 1$ 上一致分布函数 G 的经验分布函数，对于所有的实数 t，$G(t) := \max(0, \min(t, 1))$. 如果 F 是 $\mathbb{R}$ 上的任意连续分布函数，那么 $n^{1/2}(G_n(F(s)) - F(s))$ 对于它的上确界或者绝对值的上确界有同样的分布，如同对于 $G[0, 1]$ 上的一致分布一样，因为 F 为从 $\mathbb{R}$ 映上到 $(0, 1)$（F 在其值域内可能为 0 或 1，也可能不等于 0 或 1，但是在任何情况下都有 $G_n(0) - 0 = 0 = G_n(1) - 1$a. s..）同样，对于 $FG_n \circ F$ 与经验分布函数有相同的分布，这已在引理 11.4.3 中证明. 所以当 n 充分大时，从另外一个未知分布函数 F 给出 n 个独立的观测值，并且因此可以给出 F_n，我们给出检验假设“通过取 $n^{1/2}(F_n - H)$，F 是一个给定的分布函数 H”，并且求它的上确界或绝对值的上确界（这些称为“Kolmogorov-Smirnov 统计”），且确定在布朗桥的情况下对于大的或更大的值是否存在一个小概率（如 $\leqslant 0.05$），在这种情况下，不假设 $F = H$.

460

如在样本连续(定理12.1.5)的证明中所述，如果 x 是一个布朗过程，那么由 $y_t := x_t - tx_1 (0 \leq t \leq 1)$ 给出一个布朗桥. 从 x_t 到 y_t 得到布朗桥的另外一种方法是很有用的：考虑 $x_1 = 0$，y_t 的分布是 x_t 的条件分布. 由于后者仅仅是一个事件且其概率是0，这样的条件分布不是明确定义的，但是这个想法可以由下面的极限过程实现. 对于任意的 $\varepsilon > 0$，$P(|x_1| < \varepsilon) > 0$，所以条件分布 $\mathcal{L}(\{x_t\}_{0 \leq t \leq 1} \mid |x_1| < \varepsilon)$ 是定义在 $C[0, 1]$ 上的. 通常，在 $C[0, 1]$ 上，我们有由它定义的上确界范数、度量和拓扑，对于这些我们可以定义依L收敛.

12.3.2 命题 当 $\varepsilon \downarrow 0$ 时，在 $C[0, 1]$ 上有 $\mathcal{L}(x \| x_1| < \varepsilon) \to \mathcal{L}(y)$.

证明 令 $y_t := x_t - tx_1$，那么对 $[0, 1]$ 上的所有 t，有 $Ex_1y_t = 0$，根据命题9.5.12，所有这些变量都是均值为0的联合高斯分布. 根据定理9.5.14可以得出，过程 y 独立于 x_1.

令 F 是从 $C[0, 1] \times \mathbb{R}$ 映射到 $C[0, 1]$ 的函数，F 定义为 $F(g, u)(t) := g(t) + ut (0 \leq t \leq 1)$. 那么 F 是联合连续的，因此 F 是博雷尔可测的. 由于 $x_t = y_t + tx_1$，在 $C[0, 1]$ 上 $\mathcal{L}(x)$ 是像测度 $(\mathcal{L}(y) \times N(0, 1)) \circ F^{-1}$. 令 N_ε 是 x_1 的条件分布，且 $|x_1| < \varepsilon$，令 u_ε 是与过程 y 独立的随机变量，其法则是 N_ε. 换句话说，在 $C[0, 1] \times \mathbb{R}$ 上取法则 $\mathcal{L}(y) \times N_\varepsilon$，其坐标是 (y, u_ε). 那么

$$\mathcal{L}(x \| x_1| < \varepsilon) = \mathcal{L}(F(y, u_\varepsilon)).$$

因此，对于 $C[0, 1]$ 上的任意有界连续实值函数 G，当 $\varepsilon \downarrow 0$ 时，由于 $u_\varepsilon \to 0$，G 和 F 是连续的，所以

$$E(G(x) \| x_1| < \varepsilon) = E(G(F(y, u_\varepsilon))) \to E(G(F(y, 0))) = E(G(y)). \qquad \square$$

在命题12.3.2中，布朗桥 y 有时称为"固连布朗运动"，"固定在0和1的布朗运动"，或者"固定布朗运动". 另外，对于常数 c，$x_t + c$ 称为"在 c 点开始的布朗运动"(在 $t = 0$).

下一个事实是布朗桥的反射原理. y_t 达到高度 $b > 0$ 的概率是多少？它是给定 $x_1 = 0$，$x_t (t \leq 1)$ 的条件分布. 根据反射原理，时刻1时"首中"b 后到达0的概率等于在时刻1时"首中"b 后到达的 $2b$ 的概率. 但是，根据连续性，如果 $x_1 = 2b$，那么 x_t 在更早的某个时刻达到 b，所以条件概率是 $N(0, 1)$ 在 $2b$ 的密度和在0的密度的比，即 $\exp(-2b^2)$. 下面将给出详细的证明. [461]

12.3.3 命题 如果 y 是一个布朗桥，$b \in \mathbb{R}$，那么 $P(y_t = b, t \in [0, 1]) = \exp(-2b^2)$.

证明 令 $Q_b := P(y_t = b, t \in [0, 1])$. 由于在 $C[0, 1]$ 上 $\mathcal{L}(-y) = \mathcal{L}(y)$，故有 $Q_0 = 1$ 和 $Q_{-b} = Q_b$. 所以可以假设 $b > 0$. 令 $x_t := y_t + tx_1$，这里 x_1 有法则 $N(0, 1)$，且它与 y 独立. 那么 x 是一个布朗运动. 对于任意的 $\varepsilon > 0$，令

$$P_{b,\varepsilon} = P(x_t \geq b, \text{对某个 } t \in [0,1] \| x_1| < \varepsilon).$$

由于 ε 将收敛到0，故可以假设 $0 < \varepsilon < b$，那么有 $P_{b-\varepsilon,\varepsilon} \geq Q_b \geq P_{b+\varepsilon,\varepsilon}$.

令 $\tau := \inf\{t: x_t = b\}$. 那么如同在12.3.1中所证明的，$\tau$ 是一个停时. 令 $B := \{\tau < 1\}$. 根据样本连续性和中值定理知，$P(B) > 0$. 对于每一个可测事件 A，令 $P_b(A) := P_{|B}(A) := P(A \mid B)$. 那么关于 P_b，根据强马尔可夫性(定理12.2.7)可知，过程 $W_s := x_{\tau+s} - x_\tau$ 有布朗运动的法则，且与 $\mathcal{B}^B_{\tau+} := \{D \cap B: D \in \mathcal{B}_{\tau+}\}$ 独立. 根据样本连续性知，在 B 上有 $x_\tau \equiv b$. 令 $s := 1 - \tau$. 那么根据定理12.2.5知，s 是 $\mathcal{B}^B_{\tau+}$ 可测的随机变量，所以对于 P_b，它独立于 W. 如12.3.1的证明知，$\omega \mapsto W_{s(\omega)}(\omega)$ 是可测的. 现在对于 $0 < \varepsilon < b$，有

$$\begin{aligned} P_{b,\varepsilon} &= P\{\tau < 1, |W_s + b| < \varepsilon\}/P(|x_1| < \varepsilon) \\ &= P_b\{|W_s + b| < \varepsilon\}P(\tau < 1)/P(|x_1| < \varepsilon) \end{aligned}$$

(由于对任意事件 A 和 $B:=\{\tau<1\}$, $P(A\cap B)=P(A\mid B)P(B)$).

正如在 12.3.1 的证明中那样，可以记 $s=s(u)$，记 $W_t(\omega)(t\geqslant 0)$ 为 $W_t(v)(t\geqslant 0)$，其中对于 P_b，u 和 v 是独立的. 所以有 $W_{s(\omega)}(\omega)\equiv W_{s(u)}(v)$.

对于所有的 $t\geqslant 0$，用 $-W_t(v)$ 代替 $W_t(v)$ 不会改变它的分布，仍然是一个布朗运动的分布. 根据独立性知，$(s(u),\ \{W_t(v)\}_{t\geqslant 0})$ 对于 P_b(一个积测度)的联合分布等于 $(s(u),\ \{-W_t(v)\}_{t\geqslant 0})$ 的联合分布. 所以 W_s 和 $-W_s$ 对于 P_b 的分布是相同的. 因此，

$$P_b(|W_s+b|<\varepsilon)=P_b(|-W_s-b|<\varepsilon)=P_b(|W_s-b|<\varepsilon).$$

462 (关于这个反射，见图 12-3B).

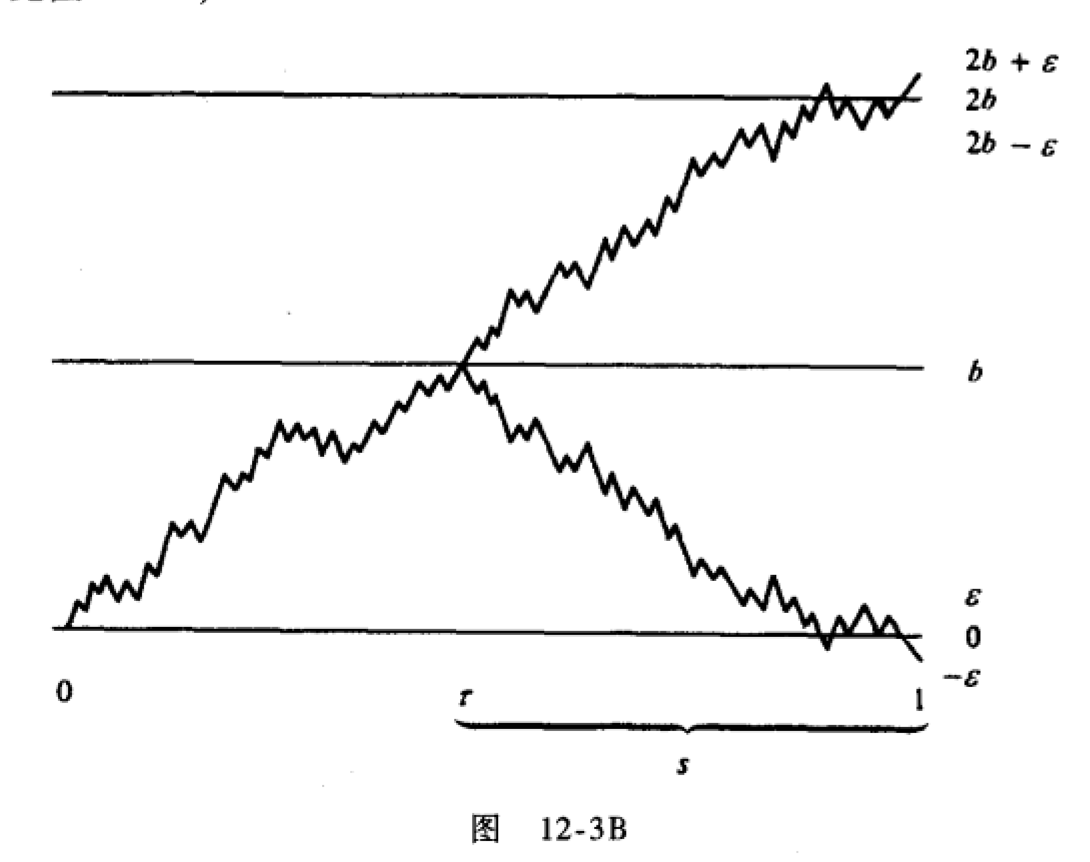


图 12-3B

根据 W_t 的定义，且由于 $x_\tau=b$，当 $\tau<1$ 时，$W_s-b=x_1-2b$. 根据样本连续性，且由于 $0<\varepsilon<b$，$|x_1-2b|<\varepsilon$ 可以推出 $x_1>b$，所以 $\tau<1$. 这可以推出

$$P_b(|W_s+b|<\varepsilon)=P(|x_1-2b|<\varepsilon)/P(\tau<1).$$

因此对于 $P_{b,\varepsilon}$，根据最后的方程，当 $\varepsilon\downarrow 0$ 时，

$$P_{b,\varepsilon}=P(|x_1-2b|<\varepsilon)/P(|x_1|<\varepsilon)\to\exp(-2b^2).$$

那么对于任意的 $\delta>\varepsilon>0$，当 ε 充分小时，

$$Q_b\leqslant P_{b-\varepsilon,\varepsilon}\leqslant P_{b-\delta,\varepsilon}<\exp(-2(b-\delta)^2)+\delta,$$

同理有

$$Q_b\geqslant P_{b+\varepsilon,\varepsilon}\geqslant P_{b+\delta,\varepsilon}>\exp(-2(b+\delta)^2)-\delta.$$

令 $\delta\downarrow 0$，即可得出 $Q_b=\exp(-2b^2)$.

□

下面研究另一个问题：求 $|y_t|$ $(0\leqslant t\leqslant 1)$ 上确界的分布. 这个问题比较难. $|y_t|$ 达到 b 的概率就是 y_t 达到 b 或 $-b$ 的概率，或者是 y_t 达到 b 的概率与它达到 b 和 $-b$ 的概率差的 2 倍. 根据命题 12.3.2，考虑 x_t 达到 b 的概率，达到 $-b$ 的概率，使得 $|x_1|<\varepsilon$ 的概率. 这等于 x_t 达到 b 的概率，达到 $3b$ 的概率，使得 $|x_1-4b|<\varepsilon$ 的概率. 最后的不等式表明：对于 $\varepsilon<b$，x_t 一定达到 b 和

463 $3b$，所以后面的概率正是 $|x_1-4b|<\varepsilon$ 的概率. 过程可能首先达到 $-b$，然后达到 b，那么在时刻

1接近于0. 正如我们反复地把一个并的概率记为概率的和减去在每一个交上的概率，在返回0之前，这些交不断地在 b 和 $-b$ 之间出现. 反射后，这些路径变成了在1时达到 b 的越来越大的倍数的路径. 最后，我们得到下面的级数.

12.3.4 命题 对于布朗桥 y 和任意的 $b>0$，

$$P(\sup\{|y_t|:0\leqslant t\leqslant 1\}\geqslant b)=2\sum_{n\geqslant 1}(-1)^{n-1}\exp(-2n^2b^2).$$

证明 令 A_n 是对于满足 $0<t_1<\cdots<t_n<1$ 的 t_j，使得 $y(t_j,\omega)=(-1)^{j-1}b(j=1,\cdots,n)$ 成立的事件. 令 $P_n:=P(A_n)$，$s:=s(\omega):=\inf\{t: y_t=b\}$，$s':=s'(\omega):=\inf\{t: y_t=-b\}$，如果 t 对应的集合是空集，其中的 s 和 s' 定义为 $+\infty$. 令 $Q_n:=P(A_n, s<s')$. 由于当用 $-y$ 代替 y 时，Q_{n+1} 不变，所以 $Q_n=P_n-Q_{n+1}$. 根据命题12.3.3知，$P_1=\exp(-2b^2)$. 根据另一个反射，如上面所表明的，有 $P_2=\exp(-8b^2)$，且迭代这个过程，则有 $P_n=\exp(-2n^2b^2)$，那么当 $n\to\infty$ 时，$Q_n\to 0$，且

$$P(\sup\{|y_t|:0\leqslant t\leqslant 1\}\geqslant b)=2Q_1=2(P_1-P_2+P_3-\cdots),$$

这就得出了所需的和. □

注：虽然和中有无穷多项，但它是交错的，且收敛速度很快，因此对于要求的任意精确度，它都很容易计算，除了在 b 值很小的时候.

另外，将会求得 $\sup_t y_t$ 和 $\inf_t y_t$ 的联合分布. 如果 $a>0$ 和 $b>0$，令

$$\Phi(a,b):=P(-a<y_t<b, 0\leqslant t\leqslant 1),$$

$$\Pi(a,b):=P(\inf_t y_t\leqslant -a, \sup_t y_t\geqslant b).$$

那么由命题12.3.3可以得出，

$$\Phi(a,b)=1-\exp(-2a^2)-\exp(-2b^2)+\Pi(a,b).$$

464

12.3.5 命题 对于任意的 $a>0$ 和 $b>0$，

$$\Phi(a,b)=\sum_{m=-\infty}^{\infty}\exp(-2m^2(a+b)^2)-\exp(-2[(m+1)a+mb]^2).$$

证明 令 A_n 是一个事件，使得对于某个 $t_1<t_2<\cdots<t_n<1$，当 j 是奇数时，$y(t_j)=-a$，当 j 是偶数时，$y(t_j)=b$. 令 B_n 是在 A_n 的定义中奇数和偶数交换后的事件. 令 $u:=u(\omega):=\inf\{t: y_t=-a\}$ 和 $v:=v(\omega):=\inf\{t: y_t=b\}$，如果 t 对应的集合是空集，则分别有 $u=+\infty$ 或 $v=+\infty$. 令 $P_n:=P(A_n)$，$Q_n:=P(A_n, u<v)$，$R_n:=P(B_n)$ 和 $S_n:=P(B_n, v<u)$. 那么 $Q_n=P_n-S_{n+1}$ 和 $S_n=R_n-Q_{n+1}$. 如同在上面两个的证明中那样，根据反射法，有

$$P_{2n}=R_{2n}=\exp(-2n^2(a+b)^2),$$

$$P_{2n+1}=\exp(-2[(n+1)a+nb]^2),$$

$$R_{2n+1}=\exp(-2[na+(n+1)b]^2).$$

因此，

$$\Phi(a,b)=1-Q_1-S_1=1+\sum_{m\geqslant 1}(-1)^m(P_m+R_m)$$

$$=\sum_{m=-\infty}^{\infty}\exp(-2m^2(a+b)^2)-\sum_{m=-\infty}^{\infty}\exp(-2[(m+1)a+mb]^2).$$ □

注：命题12.3.5可以推出命题12.3.4（对于 $a=b$），12.3.3（令 $a\to\infty$）是它的推论.

12.3.6 命题 如果 y 是一个布朗桥，那么对于任意的 $x>0$，

$$P(\sup y-\inf y\geqslant x)=2\sum_{m\geqslant 1}(4m^2x^2-1)\exp(-2m^2x^2).$$

证明 如果在命题 12.3.5 中，$\Phi(a,u)$的级数是关于u的逐项微分，则可以得到级数

$$h(a,u):=\sum_{m=-\infty}^{\infty}-4m^2(a+u)\exp(-2m^2(a+u)^2)$$
$$+4m[(m+1)a+mu]\exp(-2[(m+1)a+mu]^2).$$

它将表明对于$a+u\geqslant\varepsilon$，$a>0$和$u>0$，对于任意的$\varepsilon>0$，这个级数是一致绝对收敛和绝对收敛的. 令$v:=a+u$. 那么当$v>0$时$v\cdot\exp(-2m^2v^2)$仅在$v=1/(2|m|)<\varepsilon(|m|>1/(2\varepsilon))$时有相对最大值，所以它在$v\geqslant\varepsilon$时的值由在$v=\varepsilon$时的值控制. 我们有$\sum_m m^2\exp(-2m^2\varepsilon^2)<\varepsilon$. 当$a\geqslant0$时，$|a+mv|\exp(-2(a+mv)^2)$有最大值，要么在$a=0$时取得，这可归约为刚刚讨论的情况；要么根据微分性，如果$m\geqslant0$，当$1/2=a+mv\geqslant m\varepsilon$时，对于$m$比较大$(m>1/(2\varepsilon))$，这种情况是有可能的；要么对于$m<0$，取$1/2=|(m+1)v-u|\geqslant|m+1|\varepsilon$，对于$|m|>1+1/(2\varepsilon)$，这是不可能的. 所以一致绝对收敛成立.

465

同样，当$a\downarrow0$或$b\downarrow0$，其他保持不变时，根据命题 12.3.3 知，由于

$$\Phi(b,a)=\Phi(a,b)\leqslant\Phi(a,+\infty)=1-\exp(-2a^2),$$

所以$\Phi(a,b)\to0$. 那么对于所有的$a>0$，由于一致绝对收敛性保证了逐项可积，所以

12.3.7
$$\Phi(a,b)=\int_0^b h(a,u)\mathrm{d}u$$

令$\xi:=\inf_{0\leqslant t\leqslant1}y_t$和$\eta:=\sup_{0\leqslant t\leqslant1}y_t$，所以$\Phi(a,b)\equiv P(\xi>-a,\eta<b)$. 由 12.3.3 知，$\eta$在半直线$b>0$上的密度为$4b\cdot\exp(-2b^2)$. 对于任意的$a>0$和$u>0$，令$g(a,u):=h(a,u)/(4u\cdot\exp(-2u^2))$. 条件概率将被用到：对于每一个$a>0$和几乎所有的$\eta$，

12.3.8
$$P(\xi>-a\mid\eta)=g(a,\eta).$$

为证明这个等式，首先对于一个给定的a值，令B是在其上η可测的最小σ-代数上的一个事件，所以对于某个博雷尔集$C\subset\mathbb{R}$，$B=\eta^{-1}(C)$. 我们需要证明

12.3.9
$$P(B\cap\{\xi>-a\})=E(1_Bg(a,\eta))=E(1_C(\eta)g(a,\eta))$$

其中后一个等式是显然的. 根据式 12.3.7 知，在式 12.3.9 中的第一个等式对于$C=[0,b)(b>0)$是成立的. 因此根据微分性知，它对于$C=[c,b](c\leqslant b)$是成立的，且根据可加性知，对于这些区间的有限不交并是成立的. 根据命题 3.2.1 和命题 3.2.3 知，这些并形成了一个环. 使式 12.3.9 成立的所有C的集族是一个单调类，并且包含其中的每一个集合的补集，所以根据命题 3.2.5 和定理 4.4.2 知，它是一个σ-代数，因此它包含所有博雷尔集，即证得式 12.3.9，从而证得对于每一个$a>0$，式 12.3.8 成立. 对于几乎所有的η，

$$P(\xi\leqslant-a\mid\eta)=1-g(a,\eta).$$

对任意$x>0$，如果$\eta\geqslant x$，则$P(\eta-\xi\geqslant x\mid\eta)=1$ a.s.，所以

$$P(\eta-\xi\geqslant x)=P(\eta\geqslant x)+P(\eta<x,\xi\leqslant\eta-x).$$

假设式 12.3.8 对于所有的$a>0$和$\eta>0$成立(不仅是几乎所有的η)，这使得它对于变量$a=x-\eta$似乎是成立的. 这里是一个证明. 对于每一个$n=1,2,\cdots$把区间$0\leqslant\eta<x$分解成n个相等的区间$I_j:=[\eta_{j-1},\eta_j)$，$j=1,\cdots,n$，其中$\eta_j:=\eta(j):=jx/n$，$j=0,1,\cdots,n$. 定义事件$A_x:=\{\eta<x,\xi\leqslant\eta-x\}$，

466

$$D_n := \{对某个 j, \eta \in I_j 且 \xi \leqslant \eta_j - x\},$$
$$E_n := \{对某个 j, \eta \in I_j 且 \xi \leqslant \eta_{j-1} - x\},$$

对每个 n，显然有 $E_n \subset A_x \subset D_n$. 所以对每个 $j=1, \cdots, n$，有

$$P(D_n) = \sum_{j=1}^{n} P(\eta \in I_j, \xi \leqslant \eta_j - x)$$
$$= \sum_{j=1}^{n} \int_{\eta(j-1)}^{\eta(j)} (1 - g(x - \eta_j, u) 4u \cdot \exp(-2u^2) \mathrm{d}u,$$
$$P(E_n) = \sum_{j=1}^{n} P(\eta \in I_j, \xi \leqslant \eta_{j-1} - x)$$
$$= \sum_{j=1}^{n} \int_{\eta(j-1)}^{\eta(j)} (1 - g(x - \eta_{j-1}, u) 4u \cdot \exp(-2u^2) \mathrm{d}u.$$

令 $n \to \infty$，且注意到根据式 12.3.8 知，$g(t, u)$ 是有界的且在 $t>0$ 时是连续的(由于 $h(t, u)$ 是连续的)，则可以得到

$$P(A_x) = \int_0^x 4u(1 - g(x-u, u)) \exp(-2u^2) \mathrm{d}u.$$

那么运用命题 12.3.3 得，

$$P(\eta - \xi \geqslant x) = \exp(-2x^2) + \int_0^x 4u \cdot \exp(-2u^2) - h(x-u, u) \mathrm{d}u$$
$$= 1 + 8\sum_{m=1}^{\infty} m^2 x^2 e^{-2m^2x^2} - \sum_{m=-\infty}^{\infty} m[\mathrm{e}^{-2m^2x^2} - \mathrm{e}^{-2(m+1)^2x^2}]$$
$$= 1 - 1 + \sum_{m=1}^{\infty} [8m^2x^2 + (m-1) - (m+1)] \exp(-2m^2x^2)$$
$$= \sum_{m=1}^{\infty} (8m^2x^2 - 2) \exp(-2m^2x^2)$$

□

习题

1. 对于 $b=2$ 和 3 计算 $P(\sup_t |y_t| > b)$.
2. 对于 $x=2$ 和 4 计算 $P(\sup_s y_s - \inf_t y_t > x)$.

467

3. 令 C 是平面上通常的单位圆. 在 C 上令 P 是正规化的弧长测度 $\mathrm{d}\theta/(2\pi)$. 令 $\mathcal{A}$ 是 C 的所有子弧的集合. 证明：存在一个高斯过程 G，满足 $T=\mathcal{A}$ 且其均值是 0，这个过程使得
$$EG(A)G(B) = P(A \cap B) - P(A)P(B)$$
对于任意的弧 A 和 B 成立. 证明：G 的法则由 C 的旋转保持. 证明 G 可以由一个布朗桥 $\{y_t\}_{0\leqslant t\leqslant 1}$ 来定义，其中对于每一个形如 $\{\theta: \phi < \theta \leqslant \eta\}$ $(0 \leqslant \phi \leqslant \eta \leqslant 2\pi)$ 的弧 A，$G(A) = y_{\eta/2\pi} - y_{\phi/2\pi}$，或者对于其补弧 A^c，$G(A^c) = -G(A)$. 由 $d(A, B) := P(A\Delta B)$ 度量化 $\mathcal{A}$，其中 Δ 是对称差分，证明 G 可以是样本连续的. 证明 $\sup_{A\in\mathcal{A}} G(A)$ 有 (sup-inf)y 的分布.
4. 对于每一个实数 $v \neq 0$，求随机变量(首中时) $h_v = \inf\{t: x_t = v\}$ 对于布朗运动 x_t 的分布函数和密度函数. (证明 h_v 是几乎必然有限的). 证明：对于某常数 c 和 p，证明：当 $t \to \infty$ 时，密度 $f(t)$ 渐近于 c/t^p，并求 c 和 p 的值. [提示：运用 12.3.1 并关于 b 求微分.]
5. 对于布朗运动 x 和布朗桥 y，证明：过程 $\{x_t\}_{t\geqslant 0}$ 与 $(1+t)y_{g(t)}$ 有同样的法则，其中 $g(t) := t/(1+t)$.
6. 对于一个布朗过程 x：$x_t(\omega) := x(t)(\omega)$ 以及常数 a 和 b，求 $V(a, b) := P(x_t \geqslant at + b, t \geqslant 0)$，[提示：显然，

对于 $b\leqslant 0$，$V(a, b)=1$. 令 $b>0$. 为证明对于 $a<0$，$V(a, b)=1$，考虑足够大的 t. 对于 $a=0$，运用习题 4 的结论. 因此取 $a>0$ 和 $b>0$. 对于 $c>0$，注意到 $\{x(c^2t)/c\}_{t\geqslant 0}$ 与 $\{x_t\}_{t\geqslant 0}$ 有相同的法则，那么 $V(a, b)$ 是 ab 的函数. 在 $a=b$ 的情况下，运用习题 5 和本节的事实. 因此可以求得 $V(a, b)\equiv \mathrm{e}^{-2ab}$.]

7. 由命题 12.3.5 更详细地证明命题 12.3.3.
8. 令 $\rho\leqslant\tau$ 是布朗运动 $(x_t, \mathcal{B}_t)$ 的两个有界马尔可夫时，且对于某个常数 M，有 $\tau\leqslant M<\infty$ a. s.. 证明：$Ex_\tau^2<\infty$ 且 $E(x_\tau \mid \mathcal{B}_{\rho+})=x_\rho$ a. s.. [提示：见定理 10.4.1，把上面的 ρ 和 τ 用引理 12.2.9 的证明中的离散值停时来逼近.]
9. 在习题 8 的结论的基础上，令 $c_n<\infty$，令 $\tau(n)$ 是一个布朗运动 $(x_t, \mathcal{B}_t)$ 的马尔可夫时，且对于所有的 n 和 ω，$\tau(n)\leqslant c_n$ 和 $\tau(n)\leqslant\tau(n+1)$.

   (a) 证明：$\{x_{\tau(n)}, \mathcal{B}_{\tau(n)+}\}_{n\geqslant 1}$ 是一个鞅.

   468

   (b) 证明：对于所有的 n，$E(x_{\tau(n+1)}-x_{\tau(n)})x_{\tau(n)}=0$.
10. 证明：$\xi:=\inf_{0\leqslant t\leqslant 1} y_t$ 和 $\eta:=\sup_{0\leqslant t\leqslant 1} y_t$ 有一个联合密度 $f(v, u)$，所以对于任意的博雷尔集 $B\subset\mathbb{R}^2$，$P((\xi, \eta)\in B)=\iint_B f(v,u)\,\mathrm{d}v\mathrm{d}u$. [提示：证明对于任意的 $\delta>0$，命题 12.3.5 中的级数和它的逐项偏导数 $\partial/\partial a$，$\partial/\partial b$ 和 $\partial^2/\partial a\partial b$ 对于 $a+b\geqslant\delta>0$，$a>0$ 和 $b>0$ 是绝对一致收敛的.]
11. 对于布朗运动 x_t，令 $X:=\inf_{0\leqslant t\leqslant 1} x_t$ 和 $Y:=\sup_{0\leqslant t\leqslant 1} x_t$，对于 $a>0$ 和 $b>0$，求 $P(X\geqslant -a, Y\leqslant b)$ 的值. [提示：除了这个过程在时刻 1 处不返回到 0 外，本题类似于命题 12.3.5.]
12. 对于一个固定的 $v\neq 0$ 和在习题 4 中定义的首中时 h_v，对于 $t\neq h_v$ 和 $u(h_v, \omega):=0$，令 $u(t, \omega):=x_t(\omega)$. 证明：过程 u 和布朗运动 x_t 有相同的有限维联合分布，但 u 不是样本连续的.

12.4 在马尔可夫时布朗运动的法则：斯科罗霍德嵌入

本节的主要结果是定理 12.4.2，这个定理将要证明对于任意均值为 0 且有有限方差的随机变量 X，对于布朗运动存在着一个马尔可夫时 τ，满足 $\mathcal{L}(X)=\mathcal{L}(x_\tau)$. 这个结论以及用于独立同分布变量部分和序列的推广结论(定理 12.4.5)在下一节的研究中很有用，特别是当 $n\to\infty$ 时，讨论部分和的几乎确定性行为.

方程 $Ex_t=0$ 和 $Ex_t^2=t$ 将推广到 $Ex_\tau=0$ 和 $Ex_\tau^2=E\tau$，其中 τ 是满足 $E\tau<\infty$ 的任意马尔可夫时. 设想一个赌徒正在观测一个布朗运动，这个人可以在任意的马尔可夫时 τ 停止，然后赢得 x_τ(如果 $x_\tau<0$，将付 $-x_\tau$). 如果 τ 的期望有限，由于鞅具有有界的停时(定理 10.4.1 及其后的注)，那么平均得到的将是 0. 另一方面，令 $\tau:=\inf\{t: x_t=1\}$，那么根据 12.3.1 知，当 $b\to\infty$ 时，有 $\tau<\infty$ a. s.，但是 $Ex_\tau=1$，所以 $E\tau=+\infty$，这将在下面的定理中证明. (这个赌徒只有等才能得到钱，平均上是一个很长的时间).

12.4.1 定理 令 $(x_t, \mathcal{A}_t)_{t\geqslant 0}$ 是一个布朗运动，τ 是这个布朗运动的一个马尔可夫时，且 $E\tau<\infty$，那么 $Ex_\tau=0$ 且 $Ex_\tau^2=E\tau$.

证明 在引理 12.2.9 后已经说明了 x_τ 是可测的. 首先假设 τ 是一个停时，且它仅有有限多个值 $0\leqslant t_1\leqslant\cdots\leqslant t_n<\infty$. 对于 $n=1$，结论是显然的. 令 $t(j):=t_j$，从布朗运动 $(x_t, \mathcal{A}_t)_{t\geqslant 0}$ 的定义可以推得 $(x_{t(j)}, \mathcal{A}_{t(j)})_{1\leqslant j\leqslant n}$ 是一个鞅序列：对于 $0\leqslant s<t$，x_t-x_s 与 x_s 是独立的，所以 $E(x_t-x_s \mid x_s)=0$，因此根据可选停时(定理 10.4.1)，我们有 $Ex_\tau=Ex_{t(n)}=0$. 对 Ex_τ^2 关于 n 进行归纳是有用的：令 $\alpha:=\min(\tau, t_{n-1})$ 是一个有 $n-1$ 个值的停时. 那么

469

$$Ex_\tau^2=Ex_\alpha^2-Ex_{t_{n-1}}^2 1_{\tau=t_n}+E\left[x_{t_{n-1}}+(x_{t_n}-x_{t_{n-1}})\right]^2 1_{\tau=t_n}.$$

现在$\{\tau=t_n\}=\{\tau>t_{n-1}\}\in\mathcal{A}_{t(n-1)}$，$x_{t(n)}-x_{t(n-1)}$均值为0且根据布朗运动的定义知，它与$\mathcal{A}_{t(n-1)}$独立，所以它独立于随机变量$x_{t(n-1)}1_{\tau=t(n)}$，因此它与该变量正交．根据关于$\alpha$的归纳假设，

$$\begin{aligned}Ex_\tau^2&=E\alpha+E\left(x_{t(n)}-x_{t(n-1)}\right)^2 1_{\tau=t(n)}\\&=E\alpha+(t_n-t_{n-1})P(\tau=t_n)=E\tau.\end{aligned}$$

下面假设τ是有界的，$\tau<M<\infty$ a. s.，根据引理12.2.9后定理12.2.7的证明知，存在一个简单函数$\tau_m\downarrow\tau$且$\tau_1\leqslant M$，且这些τ_m是停时(如果$\tau_m\uparrow\tau$，它们可能不是停时)．令$\tau(m):=\tau_m$．根据样本连续性知，当$m\to\infty$时，$x_{\tau(m)}\to x_\tau$ a. s.．正如刚刚证明的，$Ex_{\tau(m)}^2=E\tau_m$，根据单调收敛性(或者控制收敛性)知，其收敛到$E\tau$.

对于每一个m和τ_m的每一个有限多个可能的值s，在$\{\tau_m=s\}$(它是在$\mathcal{A}_{\tau(m)+}$中)上，根据强马尔可夫性(定理12.2.7)知，$x_M-x_{\tau(m)}$与$\mathcal{A}_{\tau(m)+}$是条件独立的，所以有条件分布$N(0,M-s)$．因此$E(x_M\mid\mathcal{A}_{\tau(m)+})=x_{\tau(m)}$.

那么根据条件詹森不等式(10.2.7)可知，对于每个m，$x_{\tau(m)}^2\leqslant E(x_M^2\mid\mathcal{A}_{\tau(m)+})$．这可以推出$x_{\tau(m)}^2$是一致可积的(如同定理10.4.3的证明)．因此根据定理10.3.6知，$E(x_{\tau(m)}^2-x_\tau^2)\to0$，所以$Ex_\tau^2=E\tau$. 同样有$Ex_\tau=\lim\limits_{m\to\infty}Ex_{\tau(m)}=0$.

如果γ是另外一个马尔可夫时，且$\gamma\leqslant\tau\leqslant M<\infty$，那么存在简单停时$\gamma(m):=\gamma_m\downarrow\gamma$，满足$\gamma_m\leqslant\tau_m$. 由于上面证得$x_{\gamma(m)}^2$和$x_{\tau(m)}^2$是一致可积的，所以根据Cauchy-Bunyakovsky-Schwarz不等式知，$x_{\gamma(m)\tau(m)}$也是一致可积的．由于对固定的m，根据可选停时(10.4.1)知，$\{x_{\gamma(m)},x_{\tau(m)}\}$是鞅，然后运用定理10.1.9，对于所有的$m$，$E(x_{\gamma(m)}-x_{\tau(m)})x_{\gamma(m)}=0$. 所以令$m\to\infty$可得出$E(\{x_\tau-x_\gamma\})x_\gamma=0$.

现在取任意满足$E\beta<\infty$的马尔可夫时β．令$\beta(n):=\min(\beta,n)$，那么$\beta(n)$是马尔可夫时，$\beta(n)\uparrow\beta$且$Ex_{\beta(n)}^2=E\beta(n)\uparrow E\beta$. 由于$x_{\beta(n)}\to x_\beta$，由法图引理(4.3.3)可以推出$Ex_\beta^2\leqslant E\beta$. 对于$m\leqslant n$，由前面的证明可以得出$E(x_{\beta(n)}-x_{\beta(m)})x_{\beta(m)}=0$. 因此，

$$E((x_{\beta(n)}-x_{\beta(m)})^2)=E(\beta(n)-\beta(m)),$$

470

当$n\geqslant m$且$m\to\infty$时，其一致趋近于0. 再一次运用法图引理(4.3.3)得，

$$E((x_\beta-x_{\beta(m)})^2)\leqslant E\beta-E\beta(m)\to0,m\to\infty,$$

所以

$$Ex_\beta=0,\quad Ex_\beta^2=\lim_{m\to\infty}Ex_{\beta(m)}^2=\lim_{m\to\infty}E(\beta(m))=E\beta.$$

□

τ是满足$E\tau<\infty$的马尔可夫时，上面证明了$X=x_\tau$的可能法则满足$EX=0$和$EX^2<\infty$．下一个定理将证明这是X的法则上的唯一限制.

12.4.2 定理(斯科罗霍德(Skorohod)嵌入)　令$(x_t,\mathcal{A}_t)_{t\geqslant0}$是布朗运动，$\mu$是$\mathbb{R}$上的任意分布，且其满足$\int x\mathrm{d}\mu=0$和$\int x^2\mathrm{d}\mu<\infty$．那么对于$\{\mathcal{A}_t\}_{t\geqslant0}$，存在一个马尔可夫时$\tau$，$\tau$满足$\mathcal{L}(x_\tau)=\mu$和$E\tau<+\infty$，所以$E\tau=\int x^2\mathrm{d}\mu$.

注：(i)我们可以取$\mathcal{A}_t$为命题12.2.4中的最小σ-代数$\mathcal{F}_t$. (ii)如果μ集中在-1和1这两点上，那么由条件$\int x\mathrm{d}\mu=0$可以推出$\mu\{-1\}=\mu\{1\}=1/2$. 那么τ至少是满足$x_t=\pm1$的t. 另外一个例子是μ有两点支撑，这将在证明的情形I中证明.

证明　令X是法则为μ的随机变量，所以$EX=0$和$EX^2<\infty$．由定理12.4.1得出，$E\tau=$

$Ex_{\tau}^2 = EX^2$.

情形 I：假设μ集中到$-a$和b两点上，其中$a>0$和$b>0$，且$\mu\{-a\}=b/(a+b)$和$\mu\{b\}=a/(a+b)$. 令$\tau:=\inf\{t: x_t=-a \text{ 或 } b\}$. 由于$x_n-x_{n-1}(n=1, 2, \cdots)$是独立的，且其分布为$N(0, 1)$，故对所有的$n$，存在一个$\gamma>0$，使得

$$P(|x_j - x_{j-1}| \leqslant a+b \text{ for } j=1,\cdots,n) \leqslant (1-\gamma)^n.$$

如果$|x_{j+1}-x_j|>a+b$，那么要么x_j和x_{j+1}有一个比$-a$要小，要么它们之中有一个比b大. 因此$\tau<\infty$ a.s.，且由于$(1-\gamma)^n$几何地收敛到0，故$E\tau<\infty$. 由于τ是两个首中时h_{-a}和h_b的最小值，所以τ是一个停时. 在12.3.1的证明中已经证得首中时是停时. 根据定理12.2.5知，τ也是一个马尔可夫时. 由于x_{τ}仅有$-a$和b两个值，且根据定理12.4.1知，$Ex_{\tau}=0$，x_{τ}的分布是唯一确定的，所以可以得出$\mathcal{L}(X_{\tau})=\mathcal{L}(X)$.

471 情形 II：假设X是简单的，即其有有限多个可能的值(对于某个有限集F，$\mu(F)=1$).

12.4.3 引理 对于任意满足$EX=0$的简单随机变量X，存在一个概率空间$(\Omega, \mathcal{B}, P)$，$n<\infty$，使得$X_0\equiv 0$和$\mathcal{L}(X_n)=\mathcal{L}(X)$成立的鞅$\{X_j, \mathcal{B}_j\}_{0\leqslant j\leqslant n}$，并且对于每一个$j=1, \cdots, n-1$和$X_j$的每一个值，$X_{j+1}$至多有两个可能的值.

证明 概率空间取作$(\mathbb{R}, \mathcal{B}, \mu)$，其中$\mathcal{B}$是博雷尔$\sigma$-代数，$\mu=\mathcal{L}(X)$. 令$\mathcal{B}_0$是平凡代数$\{\varnothing, \mathbb{R}\}$. 令$\mathcal{B}_1$是由$(-\infty, 0]$生成的代数. 那么$\mathcal{B}_j$是一个由左开右闭的区间生成的有限代数的递增序列，如果区间A是$\mathcal{B}_j$的原子，且A是有限的$(A=(a, b])$，那么$(a, (a+b)/2]$和$((a+b)/2, b]$是$\mathcal{B}_{j+1}$的原子. 对于每个$j\geqslant 1$，$(-\infty, -j]$、$(-j, 1-j]$、$(j-1, j]$和(j, ∞)将是$\mathcal{B}_{j+1}$的原子. 因此每个$\mathcal{B}_j$有2^j个原子，每一个原子一分为二形成$\mathcal{B}_{j+1}$的原子. 由于X是简单的，且仅有有限多个值，对于充分大的j(如$j=n$)，$\mathcal{B}_j$的每一个原子将至多包含X的一个值. 那么令X_n是$\mathbb{R}$上的恒等式. X_n几乎必然等于一个$\mathcal{B}_n$可测函数. 令$X_j:=E(X_n|\mathcal{B}_j)$，$j=0, 1, \cdots, n-1$. 那么$\{X_j, \mathcal{B}_j\}_{0\leqslant j\leqslant n}$是满足所需性质的鞅. □

现在继续定理12.4.2的证明，如果X_1仅有一个可能的值，它一定是0且$X=0$ a.s.，所以令$\tau=0$. 否则，X_1有$-a$和b两个可能的值，其中$a>0$和$b>0$. 令τ_1是使得$x_t=-a$或b的最小的t. 在情形I中，$E\tau_1<\infty$，x在时刻t有X_1的法则. 令$x(t):=x_t$.

用归纳法，假设对于原来的布朗运动，给定的马尔可夫时为ρ_j，使得$E\rho_j<\infty$且$Y_j:=x(\rho_j)$有X_j的法则，满足$\tau_1=\rho_1\leqslant\rho_2\leqslant\cdots\leqslant\rho_j$.

令$z_t:=x(t+\rho_j)-Y_j$. 令$\rho(j):=\rho_j$和$\mathcal{C}_0:=\mathcal{A}_{\rho(j)+}$. 对于$t>0$，令$\mathcal{C}_t$是包含$\mathcal{C}_0$的最小$\sigma$-代数，且使得$z_s(0<s\leqslant t)$是可测的.

对于Y_j的每一个值z，其也是X_j的值，至多存在X_{j+1}的两个可能的值，记为$c=c(z)$和$d=d(z)$，其中$c=d=z$或$c<z<d$. 这里的$c(\cdot)$和$d(\cdot)$对于$\mathcal{C}_0$是(简单)随机可测变量. 令$z:=Y_j$. 令ζ是$\{\mathcal{C}_t\}_{t\geqslant 0}$的停时，定义为使得$z_t=c(z)-z$或$d(z)-z$成立的最小的$t$. 在情形I中，在$Y_j$的每一个值$z$的条件上，有$E\zeta<\infty$和$E(z_{\zeta}|\mathcal{A}_{\rho(j)+})=0$. 令$\rho_{j+1}:=\rho_j+\zeta$. 根据下面的事实，它是$\{\mathcal{A}_t\}_{t\geqslant 0}$的一个马尔可夫时.

12.4.4 引理 令ρ是布朗运动$\{x_t, \mathcal{A}_t\}_{t\geqslant 0}$的一个马尔可夫时，且$\rho<\infty$ a.s.，令$z_t:=x_{\rho+t}-x_{\rho}$，$\mathcal{C}_0:=\mathcal{A}_{\rho+}$，令$\mathcal{C}_t$是包含$\mathcal{C}_0$的最小$\sigma$-代数，对于$\mathcal{C}_t$，$z_s(0\leqslant s\leqslant t)$是可测的. 那么$\{z_t, \mathcal{C}_t\}_{t\geqslant 0}$是一个布朗运动. 令$\zeta$是$\{\mathcal{C}_t\}_{t\geqslant 0}$的一个马尔可夫时. 那么$\rho+\zeta$是$\{\mathcal{A}_t\}_{t\geqslant 0}$的一个马尔可夫时. 472

证明 要证明$\{z_t, \mathcal{C}_t\}_{t\geqslant 0}$是一个布朗运动，根据定义知，对于每个$t$，$z_t$是$\mathcal{C}_t$可测的. 接着，根

据强马尔可夫性 12.2.7 知，$\{z_u\}_{u\geq 0}$是独立于$\mathcal{C}_0$的布朗过程．对于$t\geq 0$，令$\mathcal{F}_t$是最小σ-代数，对于$\mathcal{F}_t$，$z_s(0\leqslant s\leqslant t)$是可测的．令$\mathcal{G}_t$是最小$\sigma$-代数，对于$\mathcal{G}_t$，$z_u-z_t(u\geqslant t)$是可测的．那么由于$\mathcal{C}_0$是独立于由$\mathcal{F}_{t+}$和$\mathcal{G}_t$生成的$\sigma$-代数的，而根据命题 12.2.4 知，$\mathcal{F}_{t+}$和$\mathcal{G}_t$是独立的，因此$\mathcal{C}_0$、$\mathcal{F}_{t+}$和$\mathcal{G}_t$这三个$\sigma$-代数是联合独立的，对于任意的$A\in\mathcal{C}_0$，$B\in\mathcal{F}_{t+}$和$C\in\mathcal{G}_t$，$P(A\cap B\cap C)=P(A)P(B)P(C)$．现在$\mathcal{C}_{t+}$是由$\mathcal{C}_0$和$\mathcal{F}_{t+}$生成的．下面将要证明$\mathcal{G}_t$与$\mathcal{C}_{t+}$是独立的．由于对于任意的$A$，$C\in\mathcal{A}$和$B$，$D\in\mathcal{B}$，

$$(A\cap B)\setminus(C\cap D)=[(A\setminus C)\cap B]\cup[A\cap C\cap(B\setminus D)].$$

所以对于任意两个代数$\mathcal{A}$和$\mathcal{B}$，对于$A\in\mathcal{A}$和$B\in\mathcal{B}$，所有集合$A\cap B$的类是一个半环．因此根据命题 3.2.3 知，对于$A_i\in\mathcal{C}_0$和$B_i\in\mathcal{F}_{t+}$，所有有限不交并$\bigcup_{i=1}^{n}A_i\cap B_i$的集合$\mathcal{D}$是一个代数．对于所有的$C\in\mathcal{G}_t$，使得$P(H\cap C)=P(H)P(C)$成立的集合$H\in\mathcal{C}_{t+}$的集族是一个包含$\mathcal{D}$的单调类，因此根据定理 4.4.2 知，所有$\mathcal{C}_{t+}$的集族也是一个单调类．所以$\{z_t,\mathcal{C}_t\}_{t\geq 0}$是一个布朗运动．

现在注意到对于任意固定的$r>0$，由于对于每个$t>0$，$\{\rho+r<t\}=\{\rho<t-r\}\in\mathcal{A}_v\subset\mathcal{A}_t$，其中$v:=\max(0,t-r)$，所以$\rho+r$是一个马尔可夫时．下面将要证明$\mathcal{C}_r\subset\mathcal{A}_{(\rho+r)+}$．令$B\in\mathcal{C}_r$．要证明对于任意的$t>0$，$B\cap\{\rho+r<t\}\in\mathcal{A}_t$，由引理 12.2.9 知，只需对于博雷尔集合$C\subset\mathbb{R}$和$0\leqslant s\leqslant r$，考虑形如$\{z_s\in C\}=\{x_{\rho+s}-x_\rho\in C\}$的$B$．如果$t\leqslant r$，则得到一个空集，所以可以假设$u:=t-r>0$．现在将要证明存在停时$\rho(n)$，每一个$\rho(n)$都仅有有理值，使得(a)当$n\to\infty$时，$\rho(n)\downarrow\rho$和(b)$\rho<u$蕴涵着对于所有的$n$，$\rho(n)<u$．对于(a)，运用引理 12.2.9 后面定理 12.2.7 的证明可得到有有理值的停时$\zeta(n)\downarrow\rho$．那么令$u_k\uparrow u$是有理数且$u_0:=0<u_1<\cdots<u_n<\cdots$．对于$u_{k-1}\leqslant\rho<u_k$和每个$k$，令$\xi:=u_k$．对于$\rho\geqslant u$，令$\xi:=\zeta(1)$．由于$\zeta(1)>\rho$和$\rho(n):=\min(\xi,\zeta(n))$是满足条件(a)和(b)的停时，故$\xi$是一个停时．

对于每个n，$x_{\rho(n)+s}-x_{\rho(n)}$是可测的，并且把它限制到$\{\rho+r<t\}$上是$\mathcal{A}_t$可测的．由于可测函数的极限也是可测的(定理 4.2.2)，故把$z_s=x_{\rho+s}-x_\rho$限制在$\{\rho+r<t\}$时也是$\mathcal{A}_t$可测的，且可得$\mathcal{C}_r\subset\mathcal{A}_{(\rho+r)+}$．

取任意的$t>0$．对于每个有理数$r>0$，$\{\zeta<r\}\in\mathcal{C}_r\subset\mathcal{A}_{(\rho+r)+}$．令$A_r:=\{\zeta<r\}\cap\{\rho<t-r\}$．那么由$\mathcal{A}_{(\rho+r)+}$的定义知，$A_r\in\mathcal{A}_r$．事件$\{\rho+\zeta<t\}=\{\zeta<t-\rho\}$是所有$\{\zeta<r<t-\rho\}=A_r(r>0)$的并，所以$\{\rho+\zeta<t\}\in\mathcal{A}_t$，即证得引理 12.4.4． □

473

现在回过来证明 12.4.2，$Y_{j+1}=x(\rho_{j+1})$有X_{j+1}的法则，且满足$E\rho_{j+1}<\infty$，所以归纳可以继续．那么对于$j=n$，可以得到一个马尔可夫时$\tau=\rho_n$，τ使得$E\tau<\infty$和$\mathcal{L}(X_\tau)=\mathcal{L}(X)$成立，即完成情形 II 的证明．

现在对于一般的情形，其中X不必是简单随机变量，回忆在引理 12.4.3 中$\mathbb{R}$上的σ-代数$\mathcal{B}_j$．令X_∞是$\mathbb{R}$上的恒等式，其在博雷尔σ-代数$\mathcal{B}_\infty$上有法则$\mu=\mathcal{L}(X)$．令$X_j:=E(X_\infty\mid\mathcal{B}_j)$，$j=0,1,2,\cdots$．那么$(X_j,\mathcal{B}_j)_{0\leqslant j\leqslant\infty}$是右闭鞅，所以根据定理 10.5.1 知，当$j\to\infty$时，$X_j\to X_\infty$ a.s. 如在情形 II 中那样定义$\rho_j:=\rho(j)$和Y_j．那么如在情形 II 中证明的那样，对于每个$j=0,1,\cdots$，$Y_{j+1}-Y_j=z_\zeta$，$E(z_\zeta\mid\mathcal{A}_{\rho(j)+})=0$，且根据定理 12.2.7 知，$Y_j$对于$\mathcal{A}_{\rho(j)+}$是可测的．那么根据命题 10.3.2 可以证得，$E(Y_{j+1}\mid\mathcal{A}_{\rho(j)+})=Y_j$和$(Y_j,\mathcal{A}_{\rho(j)+})_{0\leqslant j<\infty}$是鞅．因此根据定理 10.1.9 知，对于$1\leqslant j<k$，$E((Y_k-Y_j)Y_j\mid\mathcal{A}_{\rho(j)+})=Y_j^2-Y_j^2=0$．所以$E(Y_j(Y_k-Y_j))=0$，且

$$EY_k^2=E((Y_j+Y_k-Y_j)^2)=EY_j^2+E((Y_k-Y_j)^2)\geqslant EY_j^2.$$

在条件詹森不等式(10.2.7)中令$f(x):=x^2$，则可得到对于所有的k，$EY_k^2=EX_k^2=E\rho_k$和$EX_k^2\leqslant EX^2$．

所以非减马尔可夫时 ρ_j 几乎必然有一个有限极限 τ，且满足 $E\tau \leqslant EX^2$，其中对于任意的 $t>0$，$\{\tau<t\} = \cup\left\{\bigcap_{n\geqslant 1}\{\rho_n<q\}: q\in\mathbb{Q}, q<t\right\}\in\mathcal{A}_t$，所以 τ 是一个马尔可夫时. 正如 X_k 收敛到 X_∞，Y_k 几乎必然收敛到 x_τ，所以 X_k 和 Y_k 等 L 收敛到 X 和 x_τ 的法则，其也是相等的，即证得斯科罗霍德嵌入(定理 12.4.2).

例如，令 $\mu(-2)=\mu(2)=1/4$ 和 $\mu(0)=1/2$. 求使得 $\mathcal{L}(x_\tau)=\mu$ 成立的尽可能小的马尔可夫时 τ，它可以看成在立刻停止时得到 $x_\tau=0$，如掷硬币，如果它是正面，令 $\tau=0$，否则令 $\tau=\sigma$，在最小时刻 t 有 $x_t=\pm 2$. 另外一种方法是首先达到 ± 1，然后在过程达到 0 或 ± 2 时停止. 虽然看起来先到达 ± 1 再达到 0 的方法不是很有效，但是由于 $E\tau<\infty$ (定理 12.4.1)，也可以从定理 12.4.2 后的注(i)和下面的习题 5 可以看出，在两种情况下都有 $E\tau=\int x^2\mathrm{d}\mu$.

下面把定理 12.4.2 扩张到独立同分布变量部分和的序列上. 均值为 0 和方差有限的独立同分布变量 X_i 的部分和 S_n 有变量 $x_{T(n)}$ 的法则，其中 $T(n)$ 是一个马尔可夫时，当 $n\to\infty$ 时，474 $T(n)/ET(n)$ 依概率收敛到 1，在这种情况下，它是"渐近常数". 因此，与 S_n 有相同分布的 $x_{T(n)}$ 服从正态分布，可以由 $x_{ET(n)}$ 来逼近，这个逼近将在 12.5 节中证得. 它是关于序列 S_1，S_2，…性质的中心极限定理的一个改进.

12.4.5 定理 *在定理 12.4.2 的条件下，存在独立同分布的随机变量 $\tau(j)\geqslant 0(j=1, 2, \cdots)$，使得对于 $T(0):=0$ 和 $T(n):=\tau(1)+\cdots+\tau(n)$，$n=1, 2, \cdots$，每一个 $T(n)$ 是一个马尔可夫时，$x_{T(j)}-x_{T(j-1)}$ 是独立同分布的，且其法则是 μ，$Ex^2_{T(j)}=ET(j)=j\int x^2\mathrm{d}\mu<+\infty$，其中 $j=1, 2, \cdots$.*

证明 对于 $j\geqslant 1$，递归地定义 $\tau(j)$. 仅对 $1\leqslant j\leqslant n$ 运用归纳法来证明 $T(j)$ 和 $x_{T(j)}$ 是 $\mathcal{A}_{T(n)+}$ 可测的. 对于 $j=n=1$，运用斯科罗霍德嵌入 12.4.2. 根据定理 12.2.5(b)知，马尔可夫时 $\tau(1)=T(1)$ 是 $\mathcal{A}_{T(1)+}$ 可测的，且根据强马尔可夫性定理 12.2.7 知，$x_{T(1)}$ 是 $\mathcal{A}_{T(1)+}$ 可测的.

对于递归和归纳步，给定 $T(n)$，令 $z_u:=x_{T(n)+u}-x_{T(n)}$. 那么根据布朗运动的强马尔可夫性(定理 12.2.7)知，z_u 是独立于 $\mathcal{A}_{T(n)+}$ 的布朗过程，因此根据归纳假设知，它也独立于 $T(j)$ 和 $x_{T(j)}$，其中 $j\leqslant n$.

令 $\mathcal{F}_{z,u}$ 是最小的 σ-代数，如同在命题 12.2.4 中那样，$z_s(0\leqslant s\leqslant u)$ 是可测的. 根据强马尔可夫性(定理 12.2.7)知，$\mathcal{F}_{z,u}$ 中的事件与 $\mathcal{A}_{T(n)+}$ 中的事件是独立的，根据引理 12.2.9 知，$\mathcal{A}_{T(n)+}$ 包括了 $\mathcal{A}_{T(j)+}$，其中 $j<n$. 如同引理 12.4.4 中的证明，对于每个 $t\geqslant 0$，令 $\mathcal{C}_t$ 是由 $\mathcal{C}_0=\mathcal{A}_{T(n)+}$ 和 $\mathcal{F}_{z,t}$ 生成的. 那么 $\{z_t, \mathcal{C}_t\}_{t\geqslant 0}$ 是一个布朗运动. 根据斯科罗霍德嵌入(定理 12. 4. 2)和其后面的第一个注可知，对于 $\{\mathcal{F}_{z,u}\}_{u\geqslant 0}$，存在一个马尔可夫时 $\tau(n+1)$，使得 $\mathcal{L}(\tau(n+1))(n=0, 1, 2, \cdots)$ 是相同的，$\mathcal{L}(z_{\tau(n+1)})=\mu$，$\tau(n+1)$ 和 $z_{\tau(n+1)}$ 与 $\mathcal{A}_{T(n)+}$ 是独立的. 令 $T(n+1):=T(n)+\tau(n+1)$，那么根据引理 12.4.4 知，对于 $\{x_t, \mathcal{A}_t\}_{t\geqslant 0}$，$T(n+1)$ 是一个马尔可夫时. 根据定理 12.2.5(b)知 $\mathcal{A}_{T(n+1)+}$ 是可测的，根据定理 12.2.7 知，$x_{T(n+1)}$ 也是可测的. 因此递归构造可以继续，即证得定理 12.4.5. □

习题

1. 对于任意实数 c 以及布朗过程 x_t，令 $\tau:=\inf\{t: x_t=c\}$. 运用在 12.3 节的习题 4 中所得到的 τ 的精确分布，475 在不运用定理 12.4.1 的情况下，验证对于 $c\neq 0$，$E\tau=+\infty$.
2. 令 $\sigma:=\inf\{t: x_t=2 \text{ 或 } -1\}$. 求 $E\sigma$. [提示：为此不必求 σ 的分布，而求 x_σ 的分布.]
3. 求习题 2 中 σ 的分布. [提示：见 12.3 节中的习题 11，注意到对于任意的常数 c，过程 cx_t 和 x_{c^2t} 有同样的

分布.]

4. 如果 x_t 是一个维纳(布朗)过程, $\mathcal{F}_t$是最小的 σ-代数, 对于$\mathcal{F}_t$, $x_s(0\leqslant s\leqslant t)$是可测的, 且对于所有的 $t>0$, $A\in\mathcal{F}_t$, 证明: $P(A)=0$ 或 1. [提示: 运用过程有独立增量这个事实, 并比较科尔莫戈罗夫 0-1 律(8.4.4).]
5. 令 $\mu(\{-1\})=\mu(\{1\})=1/4$ 和 $\mu(\{0\})=1/2$. 考虑下面的停时: 令 η'是使得 $x_t=\pm1/2$ 成立的最小时刻, η 是最小的 $t>\eta'$, 对于 η, $x_t=-1$, 0 或 1. 令 ρ 是使得 $x_t=1$ 或 $-1/3$ 的最小时刻 t. 如果 $x_\rho=1$, 令 $\tau=\rho$; 否则, 令 τ 是使得 $x_t=0$ 或 -1 的最小的 $t>\rho$. 假设$\mathcal{A}_0$包含了满足 $P(A)=1/2$ 的事件 A. 在 A 上令 $\xi=0$. 在 A 的补集中, 令 ξ 是使得 $x_t=\pm1$ 的最小时刻.
证明: η, τ 和 ξ 这三个停时对于 μ 满足定理 12.4.2, 且有同样的期望. (在这个意义下, ξ 没有优势.)
6. 对于布朗运动$(x_t, \mathcal{A}_t)_{t\geqslant0}$, 令 τ 是使得 $x_t=1$ 或 -1 的最小的 t. 令 σ 是使得 $x_t=\pm1/2$ 的最小的 t. 令 ρ 是使得 $x_r=1/2$ 或 $-3/2$ 的最小的 $r>0$. 求 $E\tau$, $E\sigma$ 和 $E\rho$. 并验证 $E\tau=E\sigma+E\rho$.
7. 令 τ 对于某滤过$\{\mathcal{B}_t\}_{t\geqslant0}$是一个马尔可夫时, f 是从$[0,\infty]$映射到自身的非减函数, 且对于所有的 t, $f(t)\geqslant t$, f 是右连续的, $f(t)=\lim\limits_{u\downarrow t}f(u)$, 其中 $0\leqslant t<\infty$. 证明: 对于$\{\mathcal{B}_t\}_{t\geqslant0}$, $f(\tau)$也是一个马尔可夫时. 注意: f 不必是 1-1 的.

12.5 重对数律

在均值是 0 且 $EX_1^2<\infty$ 的独立同分布随机变量 X_j 的情形中, 强大数定律 $S_n/n\to0$ a. s. 可以有很大的改进: 对于任意的 $\alpha>1/2$, $S_n/n^\alpha\to0$ a. s., 这将在下面的定理 12.5.1 中证明. 另一方面, 从中心极限定理可得, 如果 $EX_1^2>0$, 那么 $S_n/\sqrt{n}$不收敛到 0. $|S_n|$ 是$\sqrt{n}$的数量级. 如果沿着序列 S_n, 将时不时地出现 $|S_n|$ 的值比$\sqrt{n}$的数量级要大. 事实上, 将会看到由 $|S_n|/(n\log\log n)^{1/2}$是几乎必然有界的, 但不是几乎必然趋近于 0. 当 $n\to\infty$ 时, $(\log\log n)^{1/2}$趋近于∞ 的速度很慢. 即可以证得对于任意的 $\alpha>0$, $|S_n|/(n^{1/2}(\log n)^\alpha)\to0$ a. s. (习题 1). [476]

对 $t>e$, 令 $u(t):=(2t\log\log t)^{1/2}$. 本节中的两个定理称为重对数律或 log log 律.

12.5.1 定理(Hartman 和 Wintner) 令 X_1, X_2, $\cdots$是独立同分布随机变量, 且 $EX_1=0$ 和$EX_1^2=1$. 令 $S_n:=X_1+\cdots+X_n$. 那么几乎必然有

(a) $\limsup\limits_{n\to\infty}S_n/u(n)=1$.

(b) $\liminf\limits_{n\to\infty}S_n/u(n)=-1$.

证明 用 $-X_j$ 取代 X_j, 由(a)可以推出(b). 下面首先证明下面的定理:

12.5.2 定理(Khinchin) 对于任意的样本连续布朗过程 x_t, $\limsup\limits_{t\to\infty}x_t/u(t)=1$ a. s..

当 $x\to\infty$ 时, 函数 F 和 G 称为渐近的, 记为 $F\sim G$, 当且仅当$\lim\limits_{x\to\infty}F(x)/G(x)=1$.

12.5.3 引理 对于任意满足 $0<\varepsilon<1$ 的 ε,

$$\limsup_{s\to\infty}\sup\{|x_t-x_s|:s\leqslant t\leqslant s(1+\varepsilon)\}/u(s)\leqslant4\varepsilon^{1/2}\text{ a. s.}$$

证明 对 $k=1,2,\cdots$, 令 $t_k:=t(k):=(1+\varepsilon)^k$. 对 $M>0$, 令

$$E_{k,M}:=\{\omega:\sup\{|x_t-x_{t(k)}|:t_k\leqslant t<t_{k+1}\}\geqslant M\}.$$

对于每个 k, 过程 $y_h:=x_{t(k)+h}-x_{t(k)}$是一个布朗过程, 在该过程中可以运用反射原理(12.3.1), 然后运用正态尾上界(引理 12.1.6(b))得到 $P(E_{k,M})\leqslant2\exp(-M^2/(2t_k\varepsilon))$.

取任意的 $\alpha>\varepsilon^{1/2}$, 令 $M:=M(k):=\alpha u(t_k)$. 那么

$$P(E_{k,M(k)})\leqslant2\exp(-\alpha^2(\log\log t_k)/\varepsilon)=2/(k\log(1+\varepsilon))^D,$$

其中 $D:=\alpha^2/\varepsilon>1$. 因此 $\sum_k P(E_{k,M(k)})<\infty$. 所以根据博雷尔–坎泰利定理(8.3.4)知，$k$ 充分大时，几乎必然有

12.5.4 $$\sup\{|x_t-x_{t(k)}|:t_k\leqslant t\leqslant t_{k+1}\}\leqslant\alpha u(t_k).$$

当 k 充分大时，$u(t_{k+1})/u(t_k)<1+\varepsilon<2$，如果 $t_k\leqslant s\leqslant t\leqslant t_{k+1}$ 和式 12.5.4 成立，那么

477

$$|x_t-x_s|/u(s)\leqslant(|x_t-x_{t(k)}|+|x_s-x_{t(k)}|)/u(t_k)\leqslant 2\alpha.$$

如果 $t_k\leqslant s\leqslant t_{k+1}\leqslant t\leqslant s(1+\varepsilon)$ 和(12.5.4)对于 k 成立，用 $k+1$ 取代 k，那么有

$$|x_t-x_s|/u(s)\leqslant(|x_t-x_{t_{k+1}}|+|x_{t_{k+1}}-x_{t_k}|+|x_s-x_{t_k}|)/u(t_k)$$
$$\leqslant\alpha u(t_{k+1})/u(t_k)+2\alpha<4\alpha.$$

令 $\alpha\downarrow\varepsilon^{1/2}$ 即证得引理. □

定理 12.5.2 的证明 对于任意的 $t>e^e$ 和 $\delta>0$，根据引理 12.1.6(b)，

$$P(x_t\geqslant(1+\delta)u(t))=N(0,1)([(1+\delta)(2\log\log t)^{1/2},\infty])$$
$$\leqslant\exp(-(1+\delta)^2\log\log t)=(\log t)^{-B},$$

其中 $B:=(1+\delta)^2>1$. 那么如同在上一个证明中，对于 $t_k:=t(k)$，同样有

$$\limsup_{k\to\infty}x_{t(k)}/u(t_k)\leqslant 1+\delta\ \text{a. s.}$$

运用引理 12.5.3 且令 $\varepsilon\downarrow 0$ 和 $\delta\downarrow 0$，则有

12.5.5 $$\limsup_{t\to\infty}x_t/u(t)\leqslant 1.$$

现在对于下界，令 $0<\delta<1$. 令 $\gamma:=\delta/2$，并取 $T>1$ 足够大，使得

12.5.6 $$(1-\gamma)(1-T^{-1})^{1/2}-(1+\delta)T^{-1/2}>1-\delta$$

对于 $j=1,2,\cdots$，令 E_j 是事件

$$E_j:=\{x(T^j)-x(T^{j-1})>(1-\gamma)u(T^j-T^{j-1})\}.$$

那么 E_j 是独立的，且由引理 12.1.6(a)中的渐近态得，$\psi(c):=(2\pi)^{-1/2}c^{-1}\exp(-c^2/2)$，

$$P(E_j)=N(0,T^j-T^{j-1})([(1-\gamma)u(T^j-T^{j-1}),\infty))$$
$$=N(0,1)([(1-\gamma)(2\log\log(T^j-T^{j-1}))^{1/2},\infty))\sim\psi((1-\gamma)(2\log\log(T^j-T^{j-1}))^{1/2})$$
$$=(2\pi)^{-1/2}(1-\gamma)^{-1}(2\log\log(T^j-T^{j-1}))^{-1/2}\cdot\exp(-(1-\gamma)^2\log\log(T^j-T^{j-1})).$$

由于 $\log\log(T^j-T^{j-1})\leqslant\log(j\log T)$，对于某个常数 C 和 $A:=1-\gamma$，有

478

$$\liminf_{j\to\infty}P(E_j)(j\log T)^A\geqslant C>0.$$

因此，$\sum_j P(E_j)$ 是发散的. 根据独立性知，对于无限多个 j，E_j 几乎必然发生.

在式 12.5.5 中，根据对称性，可以用 $-x_t$ 代替 x_t，因此也可以用 $|x_t|$ 来代替. 所以对于足够大的 j，几乎必然有 $x(T^{j-1})\geqslant-(1+\delta)u(T^{j-1})$. 那么如果 E_j 发生，则有

$$x(T^j)\geqslant(1-\gamma)u(T^j-T^{j-1})-(1+\delta)u(T^{j-1}).$$

当 $j\to\infty$ 时，$\log\log(T^j-T^{j-1})$、$\log\log(T^j)$ 和 $\log\log(T^{j-1})$ 是互相对称的. 因此，

$$\lim_{j\to\infty}u(T^j-T^{j-1})/u(T^j)=(1-T^{-1})^{1/2},$$

$$\lim_{j\to\infty}u(T^{j-1})/u(T^j)=T^{-1/2}.$$

所以根据式 12.5.6 中 T 的选择，对于无限多个 j，有 $x(T^j)\geqslant(1-\delta)u(T^j)$. 令 $\delta\downarrow 0$，可以推出 $\limsup_{t\to\infty}x_t/u(t)\geqslant 1$ a. s.，因此根据式 12.5.5 知，

$$\lim_{t\to\infty}\sup x_t/u(t) = 1.\ \text{a. s.}$$

即证得定理 12.5.2. □

定理 12.5.1 的证明 运用和的斯科罗霍德嵌入(定理 12.4.5)，其中 $\mu=\mathcal{L}(X_1)$. 那么序列 $\{x_{T(n)}\}_{n\geq 1}$ 和 $\{S_n\}_{n\geq 1}$ 有同样的分布. 由于 $\tau(n)$ 是独立同分布的，且 $E\tau(1)=1$，所以根据强大数定律，$T(n)/n\to 1$a. s.. 因此 $u(T(n))\sim u(n)$a. s.，且 $\varepsilon\downarrow 0$，根据引理 12.5.3 知，当 $n\to\infty$ 时，$(x_{T(n)}-x_n)/u(n)\to 0$a. s. 所以只需证明 $\lim_{n\to\infty}\sup x_n/u(n)=1$a. s.. 那么根据定理 12.5.2 和引理 12.5.3 知，该上界的极限几乎必然等于 1. □

习题

1. 在定理 12.5.1 的条件下，作为它的推论，证明：对于任意的 $\alpha>0$，$\lim_{n\to\infty}\sup S_n/n^{1/2}(\log n)^{\alpha}=0$a. s..
2. 令 G_n 是分布为 $N(0,1)$ 的独立同分布变量. 证明：$\lim_{n\to\infty}\sup G_n/\sqrt{\log n}=\sqrt{2}$ a. s. [提示：运用引理 12.1.6 和博雷尔–坎泰利引理(8.3.4)，用 $\sqrt{2}\pm\delta$ 取代 $\sqrt{2}$，$\delta\downarrow 0$.]
3. 令 $t(k):=t_k:=\exp(e^k)$. 求使得 $\lim_{k\to\infty}\sup x_{t(k)}/v_k=1$a. s. 的 v_k，其中 x_t 是一个布朗过程. $v_k\sim u(t_k)$ 是否成立？[提示：$x_{t(k)}-x_{t(k-1)}(k\geq 2)$ 是独立的，且当 $k\to\infty$ 时，$x_{t(k-1)}/\sqrt{t_k}\to 0$a. s.. 对 $k\geq 2$，令 $v_k:=\sqrt{2t_k\log k}$，然后运用习题 2 的结论.]
4. 令 X_j 是分布为 $N(0,1)$ 的独立同分布. 证明：$\sum_{n\geq 3}P(|S_n|>2u(n))=+\infty$. [提示：运用引理 12.1.6.(所以，重对数律不能由博雷尔–坎泰利引理证明，即使是对正态变量也不可以).]
5. 求使得 $t(m):=10^m$、$n=t(t(m))$ 和 $\sqrt{\log\log n}\geq k(k=2,5$ 和 $10)$ 成立的最小整数 m. (对数是以 e 为底的.)
6. 求 $\lim_{n\to\infty}\sup\sup_{n\leq t<n+1}(x_t-x_n)/\sqrt{\log n}$. [提示：见习题 2 和 12.3.1.]
7. 假设 X_j 是独立同分布的，且对某个 $0<\alpha<2$，满足 $E|X_j|^{\alpha}=+\infty$. 证明：重对数律对于这些 X_j 不成立，尤其是 $\lim_{n\to\infty}\sup|S_n|/u(n)=+\infty$ a. s.. [提示：运用引理 8.3.6 和博雷尔–坎泰利引理证明 $\lim_{n\to\infty}\sup|X_n|/u(n)=+\infty$ a. s.. 如果 X_n 足够大，S_n 或 S_{n-1} 也很大.]
8. 运用斯科罗霍德嵌入的方法证明：对于均值是 0、方差是正的且有限的独立同分布实随机变量 X_i，中心极限定理成立.

479

注释

12.1 节 Kolmogorov(1933a，第 3 章，§4)对于实值随机过程证明了存在性定理(12.1.2). 这个定理也被推广到概率空间的一般的投影极限上：Bochner (1955，p. 118—120)和 Frolík(1972). 回忆在任意值域空间上，积测度总是存在的(§8.2). Andersen 和 Jessen(1948)证明了在测度空间上的某些正则性假设对于随机过程的存在是必要的，修正了 Doob(1938，p. 90—93).

Doob(1953，p. 72)阐述了具有给定的非负定协方差的高斯过程的存在性(定理 12.1.3). 这是具有给定的非负定、对称协方差矩阵(9.5.7)的有限维正态分布的存在性和科尔莫戈罗夫存在性定理的重要推论.

布朗(Robert Brown)是英国的植物学家(根据 Thompson，1959，p. 73，他发现了细胞核)，他在 1827 年宣布了一个现象：悬浮在液体中的有机或无机粒子做不规则的运动(Brown，1828). 这个运动命名为"布朗运动". Bachelier(1900)讨论了一般情况下，特别是布朗运动的情形下，时间 t 连续变化的随机过程 x_t 的理论，他的研究是这类问题实质性研究的开始. 对于 Bachelier 的工作，Félix(1970)写到"……缺乏精确的考虑，在定义中的一些错误解释了为什么，尽管有创新性，但是他的

研究没有什么真正科学上的影响."

Einstein(1905, 1906, 1926)用物质的分子理论解释了布朗运动，求得了在 0 时刻位于 x 处，t 时刻的一个布朗粒子的位置分布 $N(x, at)$，其中常数 $a>0$ 主要依赖布朗粒子和液体的参数，而与 0 时刻以前的事件无关，所以在正文中，紧随布朗运动的粒子运动过程定义为高斯过程. 根据 Ulam (1957)知，Smoluchowski(1906)做了相关的工作.

Wiener(1923)证明了样本连续的布朗运动的存在性(定理 12.1.5)，其中尤其对任意的 $\alpha\leqslant 1/2$，几乎必然存在一个 $M(\omega)$，使得 $|X_s-X_t|(\omega)\leqslant M(\omega)|s-t|^{\alpha}$，$0\leqslant s\leqslant t\leqslant 1$. 上面样本连续的更简单的证明由 P. Lévy(1939, 1948)给出.

480

对于希尔伯特空间 $H=L^2([0, 1], \lambda)$，λ 为勒贝格测度，Paley 、Wiener 和 Zygmund(1933)定义了 $L(f):=\int_0^1 f(t)\,\mathrm{d}x_t$，其中 $f\in H$ 且 x_t 是一个布朗过程. 首先如果 f 是有界变量，运用分部积分法，再运用 L 是从 H 映射到 $L^2(\Omega, P)$ 的等距同构这个事实，则可以将 $L(f)$ 的定义推广到所有的 $f\in H$上. 这显然是"随机积分"的第一个定义，因此对于一个类正态过程也是成立的(与 Dudley et al. 1972 矛盾). Paley 在 1933 年去世，当时 26 岁，死于 Banff 附近的一场滑雪事故，当时这个论文正在印刷中. 在他短暂的事业生涯中，一共出版了 35 篇论文，有几篇是与别人合作的，参见 Hardy (1934)和 Poggendorff(1979). 其他重要的作品是 Paley 和 Wiener(1934)，是傅里叶分析方面的.

维纳不仅在概率和分析方面非常有名，而且在"控制论"领域也很有影响. 在计算机科学领域之外，控制理论和通信理论都将有发展. Wiener(1953, 1956)写了自传. 他的著作被 Wiener (1976—1986)所收录. Levinson(1966)是一个传记论文集，Doob(1966)收集了维纳在概率方面的著作. 也可以见 Browder、Spanier 和 Gerstenhaber(eds.)(1966).

先获得类正态过程，然后由该过程定义布朗运动过程的方法比较容易. Kahane(1976, p. 558)把这个功劳归功于 Kakutani(1944)：但我没有找到.

Itô(1944)通过允许合适的随机被积函数，实质上推广了随机积分. 也可参见 McKean(1969).

Segal(1954, 1956, etc.)在一个一般的抽象希尔伯特空间中讨论了类正态过程，且从一个不同的角度(在对偶希尔伯特空间上的有限可加测度)进行的讨论.

k 维空间中的热扩散最早是由 Fourier(1807：见 Grattan-Guinness, 1972, p. 109—111)证得，其(近似地)服从于偏微分方程(热传导方程)

$$c\frac{\partial f}{\partial t}=\sum_{j=1}^{k}\frac{\partial^2 f}{\partial x_j^2},$$

其中 c 是依赖于导热过程(均匀)介质的性质. (关于傅里叶，也可参见 Herivel(1975).) Laplace (1809)给出了对于 $t>0$ 的一个解，当 $t=0$ 时，初始值为 $g(x)$，对于 $c=k=1$，由

$$f(x,t)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}g(x+2zt^{1/2})\mathrm{e}^{-z^2}\mathrm{d}z=\frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty}g(x-u)\exp(-u^2/(4t))\,\mathrm{d}u$$

给出，其中最后一个表达式可由简单的变量替换求得，其可以看成是 g 与 $N(0, 2t)$ 的卷积. 因此 $t=0$ 时刻在 x 点的热单元可以看作是在 $t>0$ 时散播到热密度 $N(x, 2t)$上. 对于 $k>1$ 和 $c\neq 1$，有类似的公式成立.

12.2 节　对于特殊停时(如一个点的首中时)，布朗运动的强马尔可夫性从 20 世纪 30 年代开始运用，但 Hunt(1956)第一个给出了一般的、严格的叙述和定理证明. Itô 和 McKean(1956, p. 22, 26)给出了另外一种证明.

481

12.3 节 D. André (1887)处理了投票问题：如果 A 得到了 α 张选票，B 得到了 β 张选票，$\alpha>\beta$，并且投票是随机的，A 总是领先的概率是多少？André 使用一种对称的方法，这种方法可以看作是随机游动的一个反射原理，见 Feller(1968，p. 72)．Bachelier(1939，p. 29—31)和 Lévy(1939，p. 193)叙述了布朗运动的反射原理(12.3.1)，但是显然直到 Hunt(1956)才给出基于强马尔可夫性的严格证明．Bachelier(1939，p. 32)简明的叙述了反复反射的技巧．命题 12.3.3 和命题 12.3.4 融合了 Kolmogorov(1933b)和 Smirnov(1939)的研究．Doob(1949)明确地叙述了命题 12.3.3、命题 12.3.4 和命题 12.3.5. Smirnov 生于 1900 年，于 1966 年去世．精选的著作收录在 Smirnov(1970)中．Kac、Kiefer 和 Wolfowitz(1955)证明了命题 12.3.6，其被 Kuiper(1960)再次发现且应用于对均匀的圆得到一个旋转不变的检验．

12.4 节 Skorohod(1961)建立了斯科罗霍德嵌入，假设$\mathcal{A}_0$包含了所有概率介于 0 和 1 之间集合．Root(1969)证明了即使$\mathcal{A}_0$中所有事件的概率是 0 或 1，该嵌入仍然成立．进一步的结果见 Sheu(1986)和那里的参考文献．Dubins(1968)把该嵌入扩展到了鞅 z_t(其中 Ez_t^2 可能是无限的)．

12.5 节 Khinchin(1923，1924)最早发现了对于二项变量 X_j 仅有两个值的这种特殊情形的重对数律．Kolmogorov(1929)证明了某个体有界但不一定是同分布的 X_n 的重对数律，如下：令 $EX_n:=0$，$s_n:=\left(E\sum_{1\leq j\leq n}X_j^2\right)^{1/2}$．假设 $|X_n(\omega)|\leq a_n<\infty$ a. s.，其中当 $n\to\infty$ 时，$a_n(\log\log s_n)^{1/2}/s_n\to 0$ 且 $s_n\to\infty$．那么$\limsup_{n\to\infty}S_n/u(s_n^2)=1$ a. s.．

科尔莫戈罗夫的证明非常难．Stout(1974，p. 272)写道“巨大的努力是必要的”．Loève(1977，p. 266—272)给出了一个说明，给出了下面的指数下界：“令 $c=\max_{k\leq n}|X_k|/S_n\cdots$．给定 $\gamma>0$，如果 $c=c(\gamma)$足够小且 $\varepsilon=\varepsilon(\gamma)$足够大，那么 $P(S_n/s_n>\varepsilon)>\exp(-\varepsilon^2(1+\gamma)/2)$.”Tucker(1967，p. 132)认同这一点．但是对于足够大的 ε，由于变量是有界的，左边显然是等于 0 的．(在科尔莫戈罗夫的原始证明中，以不同的、正确的序选择变量．)Stout(1974，p. 262)对于指数下界给出了一个准确的假设：“令对于每个 $1\leq i\leq n$，和某个 $n\geq 1$，有 $|X_i|\leq cs_n$ a. s.，……，存在常数 $\varepsilon(\gamma)$和 $\pi(\gamma)$，使得如果 $\varepsilon\geq\varepsilon(\gamma)$和 $\varepsilon c\leq\pi(\gamma)$，那么…….”(如果 $|X_1|\leq cs_1$，那么$c\geq 1$，这使得它不可能满足其他的假设，但是如果 n 足够大且 c 足够小，则可以满足其他假设．)

Khinchin(1933)证明了布朗运动的重对数律(定理 12.5.2)．

Hartman 和 Wintner (1941)运用科尔莫戈罗夫的结论及一个巧妙的舍尾理论证明了定理(12.5.1)．Strassen(1964)运用斯科罗霍德嵌入的证明推广了这个定理及证明．Breiman(1968)给出了另一种解释．这个证明比 Kolmogorov-Hartman-Wiener 原来的证明要短得多，并且即使不包括斯科罗霍德嵌入证明的比较(12.4 节)，这个证明也要短得多，(虽然它在这本书里没有其他的用途)它有独立的意义．如果对某个 $\delta>0$，$E|X_1|^{2+\delta}<\infty$，则存在着一个不用斯科罗霍德嵌入的简单证明(Feller，1943；Pinsky，1969)．Kostka(1973)证明了这种方法在 $\delta=0$ 时不能运用．A. de Acosta(1983)不运用斯科罗霍德嵌入给出了 Hartman-Wintner 定理的一个合理而简短的证明． 482

Strassen(1966)证明了 Hartman-Wintner 重对数律的逆：如果 X_n 是独立同分布且$\limsup_{n\to\infty}|S_t|/u(n)=1$，那么 $EX_1=0$ 和 $EX_1^2=1$．用 S_n 代替 $|S_n|$ 的“一边的”逆被 Martikainen(1980)、Rosalsky(1980)和 Pruitt(1981，定理 10.1)证明．这三个人都运用了 Kesten(1970)和 Klass(1976，1977)的结论．

Bingham(1986)给出了重对数律的一般研究．这已经被推广到不同的方向上． 483

参考文献

de Acosta, Alejandro (1983). A new proof of the Hartman-Wintner law of the iterated logarithm. *Ann. Probability 11*: 270–276.

Andersen, Erik Sparre, and Børge Jessen (1948). On the introduction of measures in infinite product sets. *Danske Vid. Selsk. Mat.-Fys. Medd. 25*, no. 4, 8 pp.

André, Désiré (1887). Solution directe du problème résolu par M. Bertrand. *Comptes Rendus Acad. Sci. Paris 105*: 436–437.

*Bachelier, Louis Jean Baptiste Alphonse (1900). Théorie de la spéculation. *Ann. Ecole Norm. Sup.* (Ser. 3) *17*: 21–86.

——— (1910). Mouvement d'un point ou d'un système matériel soumis à l'action de forces dépendant du hasard. *Comptes Rendus Acad. Sci. Paris 151*: 852–855.

——— (1939). *Les nouvelles méthodes du calcul des probabilités.* Gauthier-Villars, Paris.

Bingham, N. H. (1986). Variants on the law of the iterated logarithm. *Bull. London Math. Soc. 18*: 433–467.

Bochner, Salomon (1955). *Harmonic Analysis and the Theory of Probability.* University of Calif. Press, Berkeley and Los Angeles.

Breiman, Leo (1968). *Probability.* Addison-Wesley, Reading, Mass.

Browder, Felix, E. H. Spanier, and M. Gerstenhaber (eds.) (1966). *Norbert Wiener, 1894–1964.* Amer. Math. Soc., Providence, R.I. Also in *Bull Amer. Math. Soc. 72*, no. 1, Part II.

Brown, Robert (1828). A Brief Description of Microscopical Observations made in the Months of June, July and August 1827, on the Particles contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies, London. German transl. in *Ann. Phys. 14* (1828): 294–313.

Doob, Joseph L. (1938). Stochastic processes with an integral-valued parameter. *Trans. Amer. Math. Soc. 44*: 87–150.

——— (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist. 20*: 393–403.

——— (1953). *Stochastic Processes.* Wiley, New York.

——— (1966). Wiener's work in probability theory. In Browder et al. (1966), pp. 69–71.

Dubins, Lester (1968). On a theorem of Skorohod. *Ann. Math. Statist. 39*: 2094–2097.

Dudley, R. M., Jacob Feldman, and Lucien Le Cam (1972). Some remarks concerning priorities, in connection with our paper "On seminorms and probabilities." *Ann. Math. 95*: 585.

Einstein, Albert (1905). On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat [in German]. *Ann. Phys.* (Ser. 4) *17*: 549–560. English transl. in Einstein (1926), pp. 1–18.

——— (1906). Zur Theorie der Brownschen Bewegung. *Ann. Phys.* (Ser. 4) *19*: 371–381. Transl. in Einstein (1926).

——— (1926). *Investigations on the theory of the Brownian movement.* Ed. with notes by R. Fürth. Transl. by A. D. Cowper, E. P. Dutton, New York.

Félix, Lucienne (1970). Bachelier, Louis. *Dictionary of Scientific Biography, 1*, pp. 366–367. Scribner's, New York.

Feller, Willy (1943). The general form of the so-called law of the iterated logarithm. *Trans. Amer. Math. Soc. 54*: 373–402.

——— (1968). *An Introduction to Probability Theory and Its Applications*, 1, 3d ed. Wiley, New York.

Fourier, Joseph B. J. (1822). *Théorie analytique de la chaleur.* Gauthier-Villars, Paris.

Frolík, Zdenek (1972). Projective limits of measure spaces. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 2: 67–80. Univ. of Calif. Press, Berkeley and Los Angeles.

Grattan-Guinness, I[vor] (1972). *Joseph Fourier, 1768–1830: A survey of his life and work, based on a critical edition of his monograph on the propagation of heat, presented to the Institut de France in 1807.* M.I.T. Press, Cambridge, Mass.

Hardy, Godfrey H. (1934). Raymond Edward Alan Christopher Paley. *J. London Math. Soc. 9*: 76–80.

Hartman, Philip, and Aurel Wintner (1941). On the law of the iterated logarithm. *Amer. J. Math. 63*: 169–176.

Herivel, John (1975). *Joseph Fourier: The man and the physicist.* Clarendon Press, Oxford.

Hunt, Gilbert A. (1956), Some theorems concerning Brownian motion. *Trans. Amer. Math. Soc. 81*: 294–319.

Itô, Kiyosi (1944). Stochastic integral. *Proc. Imp. Acad. Tokyo 20*: 519–524.

———, and Henry P. McKean, Jr. (1965). *Diffusion Processes and Their Sample Paths.* Springer, New York.

Kac, Mark, Jack Kiefer, and Joseph Wolfowitz (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist. 26*: 189–211.

Kahane, Jean-Pierre (1976). Commentary on Paley et al. (1933). In Wiener (1976), *1*, pp. 558–563.

Kakutani, Shizuo (1944). On Brownian motions in n-space. *Proc. Imp. Acad. Tokyo 20*: 648–652.

Kesten, Harry (1970). The limit points of a normalized random walk. *Ann. Math. Statist. 41*: 1173–1205.

Khinchin, Alexander Yakovlevich (1923). Über dyadische Brüche. *Math. Z. 18*: 109–116.

——— (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math. 6*: 9–20.

——— (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Springer, Berlin; Chelsea, New York (1948).

Klass, Michael J. (1976, 1977). Toward a universal law of the iterated logarithm. Part I, *Z. Wahrscheinlichkeitsth. verw. Geb. 36*: 165–178, Part II, ibid. *39*: 151–165.

Kolmogorov, Andrei N. (1929). Über das Gesetz des iterierten Logarithmus. *Math. Ann. 101*: 126–135.

——— (1933a). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin. Transl. as *Foundations of Probability.* Chelsea, New York (1956).

——— (1933b). Sulla determinazione empirica di una legge di distribuzione (in Italian). *Giorn. Ist. Ital. Attuar. 4*: 83–91. Russian transl. in Kolmogorov (1986), 134–141.

——— (1986). *Probability Theory and Mathematical Statistics* (selected works; in Russian). Moscow, Nauka.

Kostka, D. G. (1973). On Khintchine's estimate for large deviations. *Ann. Probability 1*: 509–512.

Kuiper, Nicolaas H. (1960). Tests concerning random points on a circle. *Proc. Kon. Akad. Wetensch. A (Indag. Math. 22) 63*: 38–47.

Laplace, Pierre Simon de (1809). Mémoire sur divers points d'analyse. *J. Ecole Polytechnique*, cahier 15, tome 8, pp. 229–264; *Oeuvres, XIV*, pp. 178–214, esp. pp. 184–193.

Levinson, Norman (1966). Wiener's life. In Browder et al. (1966), pp. 1–32.

Lévy, Paul (1939). Sur certains processus stochastiques homogènes. *Compositio Math. 7*: 283–339.

——— (1948). *Processus stochastiques et mouvement brownien.* Gauthier-Villars, Paris.

Loève, Michel (1977). *Probability Theory 1.* 4th ed. Springer, New York.

Martikainen, A. I. (1980). A converse to the law of the iterated logarithm for a random walk. *Theory Probability Appl.* *25*: 361–362.

McKean, Henry P. (1969). *Stochastic Integrals*. Academic Press, New York.

Paley, Raymond Edward Alan Christopher, Norbert Wiener, and Antoni Zygmund (1933). Notes on random functions. *Math. Z.* *37*: 647–668. Also in Wiener (1976), *1*, pp. 536–557.

———, and N. Wiener (1934). *Fourier Transforms in the Complex Domain. Amer. Math. Soc. Colloq. Publs. 19*.

Pinsky, Mark (1969). An elementary derivation of Khintchine's estimate for large deviations. *Proc. Amer. Math. Soc.* *22*: 288–290.

J. C. Poggendorffs biographisch-literarisches Handwörterbuch 7b Teil 6 (1979). Paley, R. E. A. C. Akademie-Verlag, Berlin.

Pruitt, William E. (1981). General one-sided laws of the iterated logarithm. *Ann. Probab.* *9*: 1–48.

Root, David H. (1969). The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.* *40*: 715–718.

Rosalsky, Andrew (1980). On the converse to the iterated logarithm law. *Sankhyā* Ser. A *42*: 103–108.

Segal, Irving Ezra (1954). Abstract probability spaces and a theorem of Kolmogoroff. *Amer. J. Math.* *76*: 721–732.

——— (1956). Tensor algebras over Hilbert spaces, I. *Trans. Amer. Math. Soc.* *81*: 106–134.

Sheu, Shey Shiung (1986). Representing a distribution by stopping a Brownian motion: Root's construction. *Bull. Austral. Math. Soc.* *34*: 427–431.

Skorohod, Anatolii Vladimirovich (1961). *Studies in the Theory of Random Processes* [in Russian]. Univ. of Kiev. Transl. Addison-Wesley, Reading, Mass. (1965).

Smirnov, Nikolai Vasil'evich (1939). Estimation of the deviation between empirical distribution curves of two independent samples [in Russian]. *Bull. Univ. Moscow* *2*, no. 2, pp. 3–14. Repr. in Smirnov (1970), pp. 117–127, 267.

——— (1970). *Theory of probability and mathematical statistics: Selected works* [in Russian]. Nauka, Moscow.

Smoluchowski, Marian (1906). Zur kinetischen Theorie der Brownschen Molekular–bewegung und der Suspensionen. *Ann. Phys.* (Ser. 4) *21*: 756–780.

Stout, William F. (1974). *Almost Sure Convergence*. Academic Press, New York.

Strassen, Volker (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. verw. Geb.* *3*: 211–226.

——— (1966). A converse to the law of the iterated logarithm. *Z. Wahrsch. verw. Geb.* *4*: 265–268.

Thompson, D'Arcy (1959). *Growth and Form*. Cambridge Univ. Press. (1st ed. 1917.)

Tucker, Howard G. (1967). *A Graduate Course in Probability*. Academic Press, New York.

Ulam, Stanislaw (1957). Marian Smoluchowski and the theory of probabilities in physics. *Amer. J. Phys.* *25*: 475–481.

Wiener, Norbert (1923). Differential space. *J. Math. Phys. M.I.T.* *2*: 131–174. Also in Wiener (1976), *1*, pp. 455–498.

——— (1953). *Ex-Prodigy: My Childhood and Youth*. Simon & Schuster, New York. Repr. M.I.T. Press, Cambridge, Mass. (1964).

——— (1956). *I Am a Mathematician*. Doubleday, New York. Repr. M.I.T. Press, Cambridge, Mass. (1964).

——— (1976–1986). *Norbert Wiener: Collected Works with Commentaries*. Ed. Pesi Masani. 4 vols. M.I.T. Press, Cambridge, Mass.

第13章　可测性：博雷尔同构和解析集

*13.1　博雷尔同构

两个可测空间$(X, \mathcal{B})$和$(Y, \mathcal{C})$称为是同构的(isomorphic)，当且仅当存在一个从X映射到Y的一对一函数f，使得f和f^{-1}都是可测的．两个度量空间(X, d)和(Y, e)称为博雷尔同构的，当且仅当它们的博雷尔集的σ-代数是同构的，且具有博雷尔集的σ-代数，记作$X \sim Y$.

显然，博雷尔同构是在同胚拓扑和集合同构之间导出的，这意味着具有相同的势．下面的定理表明在许多情况下，博雷尔同构与有相同的势是等价的．

13.1.1 定理　*如果X和Y是两个可分的度量空间，它们是X和Y的完备化博雷尔子集，那么$X \sim Y$，当且仅当X和Y有相同的势，而且它们或者是有限可数的，或者为c(连续统的势，即$[0, 1]$的势).*

注：一般地，连续统假设规定没有集合具有不可数但严格小于c的势，这与集合论中包括选择公理在内的其他公理是独立的(见附录A.3的注释)．然而，对于完备可分度量空间中的博雷尔集合而言，由以下将要证明的定理可知，连续统假设服从公理．同构的例子有$\mathbb{R} \sim \mathbb{R}^{\infty}$和$\mathbb{R} \sim \mathbb{R} \setminus \mathbb{Q}$(无理数空间).

证明将以其他几个事实为基础．对于任意度量空间S，令S^{∞}是S的一个可数积，具有积拓扑，这在命题2.4.4中被度量化．如果S是完备的，那么根据定理2.5.7知，具有此度量的S^{∞}也是完备的.

令"2"表示具有两个点$\{0, 1\}$的离散空间，那么2^{∞}将会是紧度量空间，它是$\{0, 1\}$的可数积(并且，同样地，通过$f(\{x_n\}) = \sum_n 2x_n/3^n$，与命题3.4.1中讨论的康托尔集是同胚的)．通常，令$I := [0, 1]$.

13.1.2 引理　*存在满足$B \sim I$和$C \sim I^{\infty}$的博雷尔集$B \subset 2^{\infty}$和$C \subset 2^{\infty}$.*

证明　令B是满足或者对所有的n，$x_n = 1$，或者对无限多个n，$x_n = 0$的所有$\{x_n\} \in 2^{\infty}$的集合，那么B在2^{∞}中有可数补，所以它是一个博雷尔集．通过$f(\{x_n\}) = \sum_n x_n/2^n$(二项展开)定义$f$，那么$f$从$2^{\infty}$映上到$I$，并且从$B$映射到$I$是1-1的．由于定义$f$的级数一致收敛并且它们的有限部分和是连续的，所以$f$是从$2^{\infty}$映上到$I$的连续函数，从而是博雷尔可测的．另一方面，显然对$f$到$B$上的限制的逆，在$I$上每一个数字$x_n$都是可测的．从$I$映上到一个有限集$2^n$的每一个$n$元数组$(x_1, \cdots, x_n)$都是可测的．对于$x \in I$，令$g_n(x) = (x_1(x), \cdots, x_n(x), 0, 0, \cdots)$，则$g_n(x)$是可测的．那么$g_n$逐点收敛于$f^{-1}$，从而$f^{-1}$也是可测的(定理4.2.2)，因此$B \sim I$．那么$B^{\infty} \sim I^{\infty}$，并且$(2^{\infty})^{\infty} \sim 2^{\infty}$，所以对于某个博雷尔集$C \subset 2^{\infty}$，$B^{\infty} \sim C$.　□

13.1.3 引理　*对任意完备可分度量空间Y和博雷尔子集X，存在满足$X \sim A$和$Y \sim B$的博雷尔集$A \subset B \subset I^{\infty}$.*

证明　由命题2.4.4和定理2.8.2的证明知，Y同胚于I^{∞}的子集B，那么根据定理2.5.4可知，B是开集(G_δ)的一个可数交(本身闭包的交)，从而是一个博雷尔集．于是博雷尔同构$Y \sim B$给出了

一个满足 $A \sim X$ 的博雷尔集 A(由于博雷尔集的博雷尔子集还是博雷尔集，这一点可以从 A 是相对的开集等得到.) □

通常，在非负整数空间$\mathbb{N}$上有离散拓扑.

13.1.4 引理　*对任意(非空)的完备可分度量空间(X, d)，存在一个从$\mathbb{N}^\infty$映射到 X 的连续函数 f.*

例：令 $X=[0, 1]$，具有通常的度量. 对任意的整数 n，令 $L(n)$ 是 n 的最后一位十进制数字，例如 $L(317)=7$. 对任意非负整数序列$\{n_i\}_{i\geq 1}$，令$f(\{n_i\}) := \sum_i L(n_i)/10^i$ 为数字用 $L(n_i)$ 表示的

488

十进制展开，那么 f 从$\mathbb{N}^\infty$映上到$[0, 1]$是连续的.

证明　对 X 的任意子集 A，令 $\mathrm{diam}(A) := \sup\{d(x, y): x, y \in A\}$，称为 A 的直径. 那么有 $X=\bigcup_{n\geq 1} A_n$,其中每一个 A_n 都是直径 $\mathrm{diam}(A_n)\leq 1$ 的非空闭集(A_n 一般是不相交的). 令 $A(n) := A_n$，递归地，对 $k=1, 2, \cdots$存在直径至多为 $1/k$ 的非空闭集 $A(n_1, \cdots, n_k)$，使得对每一个 k 和 $n_j \in \mathbb{N}$，$j=1, \cdots, k$,

$$A(n_1,\cdots,n_k) = \bigcup_{i\in\mathbb{N}} A(n_1,\cdots,n_k,i).$$

那么对每个$\{n_j\}\in\mathbb{N}^\infty$，$\{F_k\}_{k\geq 1} := \{A(n_1, \cdots, n_k)\}_{k\geq 1}$是一个非空闭集的递减序列. 选取任意一个序列 $x_k \in F_k$，对 $k \geq m$，我们有 F_m 中的一个柯西序列，从而这个序列收敛到 $x \in \bigcap_k F_k$. 因为 $\mathrm{diam}(F_k)\downarrow 0$，这个 x 是唯一的. 令 $f(\{n_j\}_{j\geq 1})=x$.

因为$\mathbb{N}$具有离散拓扑，如果一个序列$\{z_n\}$在$\mathbb{N}^\infty$中收敛到某个 y，那么在$\mathbb{N}$中每一个坐标 z_{nk} 也收敛并且最终等于 y_k(y 的第 k 个坐标). 一旦 z_n 的前 m 个坐标被固定，$f(z_n)$也随之移动至多 $1/m$ 的距离. 因此 $f(z_n)$收敛，所以 f 是连续的. 显然 f 是满射. □

13.1.5 定理　*对完备可分度量空间 X 中的任意非空博雷尔集 B，存在一个从$\mathbb{N}^\infty$映上到 B 的连续函数 f.*

例：令 B 是有理数集$\mathbb{Q}$，由推论 2.5.6 知，对于任意度量化其通常拓扑(作为$\mathbb{R}$的子集的相对拓扑)的度量，B 都是不完备的. 令$\mathbb{Q}=\{q_n\}_{n\geq 0}$. 对任意坐标为 $n(1)$，$n(2)$，$\cdots$的 $n\in N^\infty$，令 $f(n) := q_{n(1)}$，那么 f 从$\mathbb{N}^\infty$映上到$\mathbb{Q}$是连续的.

证明　令$\mathcal{C}$是 X 中所有博雷尔集的集族，X 是$\mathbb{N}^\infty$上连续函数的值域. 那么由引理 13.1.4. 可知，所有完备的闭集都在$\mathcal{C}$中.

对 $n=1, 2, \cdots$令 $A_n\in\mathcal{C}$. 对每个 n，令 f_n 从$\mathbb{N}^\infty$映上到 A_n 是连续的. 对$\{n_j\}_{j\geq 1}\in\mathbb{N}^\infty$，令 $f(\{n_j\}) := f_{n(1)}(\{n_{j+1}\}_{j\geq 1})$，其中 $n(1) := n_1$. 那么 f 把$\mathbb{N}^\infty$映上到$\bigcup_n A_n$，且 f 是连续的，因为$\mathbb{N}$上的拓扑是离散的(对 n_1 坐标)，而且每个 f_n 都是连续的. 所以 $\bigcup_n A_n\in\mathcal{C}$. 度量空间上的开集是 F_σ 集，也就是闭集的可数并(正如定理 7.1.3 的证明中)，所以所有的开集都在$\mathcal{C}$中.

令 $F := \{\{\alpha_j\}_{j\geq 1}\in(\mathbb{N}^\infty)^\infty$：对所有的 j，$f_1(\alpha_1)=f_j(\alpha_j)\}$. 那么 F 是闭集的交，因此是闭的. 通

489

过 $g(\alpha) := f_1(\alpha_1)$在 F 上定义 g，那么 g 是连续的，并且把 F 映上到 $\bigcap_j A_j$. 由引理 13.1.4 知，存在一个从$\mathbb{N}^\infty$映上到 F 连续函数 h，所以 $g\circ h$ 从$\mathbb{N}^\infty$映上到 $\bigcup_j A_j$ 是连续的，因此也在$\mathcal{C}$中.

令$\mathcal{D}$是满足 B 和 $X\setminus B$ 在$\mathcal{C}$中的集合 B 的集族，那么所有开集都在$\mathcal{D}$中. $\mathcal{D}$中集合的任意可数并也在$\mathcal{D}$中. 如果 $B\in\mathcal{D}$，那么 $X\setminus B\in\mathcal{D}$. 因此$\mathcal{D}$是一个 σ-代数，并且与整体博雷尔 σ-代数相等，因

此也就等于C. □

定义 拓扑空间中的集合S称为**自稠密的**(dense in itself)，如果对S中的每一个x，点x的每一个邻域都包含S中除了x外的其他点. 一个自稠密紧集称为是**完全的**(perfect).

13.1.6 引理 对任意可分度量空间X，存在一个可数集$C\subset X$，使得$X\setminus C$是自稠密的.

证明 令C是所有使得X中y的某个开邻域是可数的$y\in X$的集合. 这种开邻域的集族给出了C的一个开覆盖，这个开覆盖有可数个子覆盖("Lindelöf 定理"，特别地，由命题2.1.4可以在X的拓扑的某个可数基中取邻域). 所以C是可数的，由C的定义知，它的补集也是自稠密的. □

13.1.7 定理(Alexandroff-Hausdorff) 在完备可分度量空间(X, e)中，每一个不可数博雷尔集B都包含一个同胚于2^{∞}的完全集C.

例：$[0, 1]$上的康托尔集C是所有和$x = \sum_{i\geq 1} n_i/3^i$的集合，其中对于所有的$i$，$n_i=0$或2. C是完全的，并且通过x与$\{n_i/2\}_{i\geq 1}$的对应，C与2^{∞}同胚.

证明 由定理13.1.5知，存在一个从完备可分度量空间(S, d)映上到B的连续函数f. 对每个$y\in B$，在S中选取一个满足$f(x)=y$的$x:=x_y$. 令A是所有x_y的集合. 利用引理13.1.6，通过剔除一个可数集，可以假设A是自稠密的. 在A中取出任意两个不相同的点x_0和x_1，那么因为f是连续的而且在A上是1-1的，所以在A中存在x_i的不相交闭邻域F_i，使得$f(F_i)$的值域的闭包是不相交的. 同样地，每一个F_i包括两个内部(都在A的相对拓扑中)非空的闭集F_{i0}和F_{i1}，使得值域$f(F_{ij})$的闭包是不相交的. 继续递归，对$m=1, 2, \cdots$我们得到闭集$F_{i(1)i(2)\cdots i(m)}$，其中对每个k，$i(k)=0$或1，对每个m，f在这些集合上的值域的闭包都是不相交的. 同样，对所有的m和$i(1)$，$i(2)$，$\cdots$，$i(m)$，选取满足$F_{i(1)i(2)\cdots i(m)}\subset F_{i(1)i(2)\cdots i(m-1)}$的集合. 在所有情况下，还可以假定对所有的$x$，$y\in F_{i(1)\cdots i(m)}$，$d(x, y)<1/m$.

490

对每个$i=\{i(k)\}_{k\geq 1}\in 2^{\infty}$，所有由$S$中一个唯一的点组成的$F_{i(1)\cdots i(m)}$的交，称为$g(i)$. 这给出了从$2^{\infty}$映射到$S$的一个函数. 显然，$g$是1-1的并且是连续的. 根据选择，$f$在$g$的值域上是1-1的. 因此$f\circ g$从$2^{\infty}$映射到$B$是1-1和连续的. 由于$2^{\infty}$是紧的，所以由定理2.2.11知，$f\circ g$是一个同胚. 现在根据拓扑积的定义可知，$2^{\infty}$是完全的，所以$f\circ g$的值域也是完全的. □

13.1.8 引理 如果A，B和C是完备可分度量空间S的博雷尔子集，满足$A\subset B\subset C$和$A\sim C$，那么$A\sim B$.

注：令$A\subset B\subset C$为任意集合，假设存在一个从A映上到C的1-1函数. 那么根据等价定理(1.4.1)可知，从A映上到B存在一个1-1函数. 在这种意义下，引理13.1.8的结论与波兰空间类似，并且与一般函数的可测函数等价定理类似.

证明 令$A_0:=A$，$D_0:=C/A$. 递归地，对$n=0, 1, \cdots$和不相交的博雷尔集A_n和D_n，令f_n是映上到$A_n\cup D_n$的A_n的一个博雷尔同构，所以f_0存在. 令$A_{n+1}:=f_n^{-1}(A_n)$，$D_{n+1}:=f_n^{-1}(D_n)$. 那么$A_n=A_{n+1}\cup D_{n+1}$，其中A_{n+1}和D_{n+1}是不相交的博雷尔集，经由f_n，$A_{n+1}\sim A_n$，$D_{n+1}\sim D_n$. 令f_{n+1}是把f_n限制到A_{n+1}上得到的，A_{n+1}映上到$A_{n+1}\cup D_{n+1}$的一个博雷尔同构. 所以可以继续进行递归. 令$E:=\bigcap_n A_n$，那么E，D_0，$D_1\cdots$是满足$A=E\bigcup_{n\geq 0} D_n$的不相交的博雷尔集.

令$F_0:=C\setminus B$，$G_0:=B\setminus A$. 那么对所有的n，$D_0=F_0\cup G_0$的分解在满足$F_{n+1}\sim F_n$，$G_{n+1}\sim G_n$的不相交博雷尔集上产生一个分解$D_n=F_n\cup G_n$.

如果对所有的n，X_n(在某些X中)是不相交的博雷尔集，Y_n(在某些Y中)是不相交的博雷尔

集，且满足 $X_n \sim Y_n$，那么(根据引理 4.2.4) $\bigcup_n X_n \sim \bigcup_n Y_n$. 所以

$$C = E \cup \bigcup_{n\geq 0} D_n = E \cup \bigcup_{n\geq 0} F_n \cup \bigcup_{n\geq 0} G_n,$$

$$\sim E \cup \bigcup_{n\geq 1} F_n \cup \bigcup_{n\geq 0} G_n = C \setminus F_0 = B,$$

所以 $A \sim B \sim C$. □

定理 13.1.1 的证明 显然，如果 $X \sim Y$，那么 X 和 Y 具有相同的势. 反之，如果 X 是可数的，那么它的博雷尔集的 σ-代数包括所有的子集，所以，如果 Y 也有相同的势，那么 $X \sim Y$.

491

假设 X 是不可数的，那么由定理 13.1.7 知，X 包含一个同胚于 2^ω 的集合 K. 另一方面，由引理 13.1.3 知，对某个博雷尔集 $H \subset I^\omega$，$X \sim H$，所以根据引理 13.1.2 知，对一个博雷尔集 $D \subset 2^\omega$，$X \sim H \sim D$. 所以对某个博雷尔集 A，$2^\omega \sim A \subset D \subset 2^\omega$，那么由引理 13.1.8 知，$X \subset D \subset 2^\omega$. 所以，如果 Y 也是不可数的，那么 $Y \sim 2^\omega \sim X$. □

习题

1. 证明：$[0, 1]$与 2^ω 不是同胚的. [提示：一个拓扑空间 X 称为连通的(connected)，如果它不是两个非空不相交开集的并.]
2. 拓扑空间 X 称为完全不连通的(totally disconnected)，当且仅当对 X 中任意两个点 $x \neq y$，存在不相交的开集 U 和 V，使得 $X = U \cup V$，$x \in U$，$y \in V$. 证明：2^ω 是完全不连通的.
3. 令 X，Y 是两个不可数的无限豪斯多夫拓扑空间，证明：存在一个从 X 映上到 Y 具有博雷尔逆的 1－1 博雷尔可测函数.
4. 求$[0, 1]$的一个具体的可数无限子集 A，和从$[0, 1]$映上到$[0, 1] \setminus A$ 的博雷尔同构. [提示：参考命题 12.1.1 及其证明.]
5. 从$[0, 1]$映上到 2^ω，给出一个具体的具有博雷尔逆的 1－1 博雷尔函数. [提示：参考习题 4 和引理 13.1.2 的证明.]
6. 给出一个具体的从$[0, 1]$映上到$[0, 1] \times [0, 1]$具有博雷尔逆的 1－1 博雷尔函数. [提示：利用习题 5.]
7. 证明或反证：在满足 $X \setminus C$ 是自稠密的可数集合 C 中，对每一个可分度量空间(X, d)总存在(a)一个最大集合 C，(b)一个最小集合 C. [提示：参考 1.3 节.]
8. 证明或反证：令 $X := (-2, -1) \cup \mathbb{N}$ 为具有通常拓扑的$\mathbb{R}$ 的子集. 对完备可分度量空间中的每个博雷尔集 B，存在一个从 X 映上到 B 的连续函数. [提示：考虑 $B = 2^\omega$.]
9. 令(S, d)是可分度量空间，(T, e)是一个度量空间. 令 f 是从 S 映射到 T 的一个博雷尔可测函数. 在连续统假设条件下，证明：值域 $f[S]$是可分的. [提示：如果不可分，证明 $f[S]$包含一个势为 c 且具有离散相对拓扑的不可分闭集 A. 因此，A 的所有子集 C 都是闭的，所有集合 $f^{-1}(C)$是 S 中的博雷尔集，从而 2^c 是 S 中的博雷尔集，正如 4.2 节的习题 8 所述，这是不可能的.]

492

10. 令 Ω 是最小的不可数序数，即$(\Omega, <)$是不可数的良序集，使得对每个 $\alpha \in \Omega$，$\{\beta \in \Omega: \beta < \alpha\}$ 是可数的. 在 Ω 上，取具有子基的区间拓扑，子基是由所有开区间$\{\gamma: \gamma < \beta\}$ $(\beta \in \Omega)$和$\{\gamma: \gamma > \alpha\}$ $(\alpha \in \Omega)$给出的. 证明：不存在从 Ω 映上到$[0, 1]$的 1－1 博雷尔可测函数 f. [提示：如果 f 是从 Ω 映上到 2^ω 的博雷尔可测函数，可以根据 7.3 节习题 2 的方法证明对某个 $n = \{n_j\} \in 2^\omega$，$f^{-1}\{n\}$ 是不可数的.]

13.2 解析集

到目前为止在这本书中，度量空间中的可测集一般是博雷尔集，或者对有个测度的完备化是可测集，例如，直线上的勒贝格可测集. 博雷尔 σ-代数是由空间拓扑生成的，所以对于这个拓扑，

它不依赖于具体的度量. 回想一个由一种度量度量的拓扑空间，并且对于此度量，它是完备可分的，称为波兰空间. 由命题2.4.4和定理2.5.7知，任意一个具有积拓扑的可数个波兰空间的笛卡儿积是波兰的. 定理13.1.5表明波兰空间中的任意博雷尔集都是$\mathbb{N}^\infty$的连续映像，本身还是一个波兰空间. 例如，如果V是$\mathbb{R}^k$中的开集或闭集，那么V是紧集的可数并，所以它是本身的任意连续像，特别地，也是一个博雷尔集. 令人惊奇的是，每一个$\mathbb{N}^\infty$的连续像不是博雷尔集. 博雷尔集的连续或博雷尔可测的像描述如下. 回想由$f[A]:=\{f(x): x\in A\}$所定义的函数f作用在集合A上的直接像$f[A]$.

13.2.1 定理 令Y是一个波兰空间，A是Y的非空子集，那么下面6个条件是等价的：

(a) 对某个连续函数f，$A=f[\mathbb{N}^\infty]$.

(a′) 对某个博雷尔可测函数f，$A=f[\mathbb{N}^\infty]$.

(b) 对某个波兰空间X和连续函数f，$A=f[X]$.

(b′) 对某个波兰空间X和博雷尔可测函数f，$A=f[X]$.

(c) 对波兰空间X中的某个博雷尔集B和从B映射到Y的连续函数f，$A=f[B]$.

(c′) 对波兰空间X中的某个博雷尔集B和从B映射到Y的博雷尔可测函数f，$A=f[B]$.

证明 由于任意连续函数f是博雷尔可测函数(由定理4.1.6)，因此(a)⇒(a′)，(b)⇒(b′)，(c)⇒(c′).

493

由于$\mathbb{N}^\infty$是波兰的(由命题2.4.4和定理2.5.7)，故(a)⇒(b)，(a′)⇒(b′). 显然(b)⇒(c)，(b′)⇒(c′)，所以只需证明(c′)⇒(a). 在(c′)中，对所有的$x\in X\setminus B$，选择$c\in A$，令$f(x):=c$，所以(b′)成立，我们需要证明(b′)⇒(a).

现在$X\times Y$是波兰空间. 假设在$X\times Y$中f的图形H是博雷尔可测的. 例如，连续函数的图形总是闭的. 令g是从$X\times Y$映上到Y的投影$g(x, y):=y$. 那么g是连续的. 由定理13.1.5知，对某个连续函数h，$H=h[\mathbb{N}^\infty]$，那么

$$A=(g\circ h)[\mathbb{N}^\infty]=g[h[\mathbb{N}^\infty]],$$

其中$g\circ h$是连续的，证明了(a). 所以我们只需证明下面的引理.

13.2.2 引理 如果X，Y是波兰空间，f是从X映上到Y的一个博雷尔可测函数，那么f的图形在$X\times Y$中是博雷尔集.

证明 由博雷尔同构定理(13.1.1)可知，Y与2^∞或者2^∞的某个可数子集是博雷尔同构的，所以可以假设$Y=2^\infty$. 对每个$n=1, 2, \cdots$令

$$T(n):=\{s=\{s_j\}_{1\leqslant j\leqslant n}: \text{对每个}\ j,\ s_j=0\ \text{或}\ 1\}.$$

对每个$s\in T(n)$，令

$$C_s:=\{u=\{u_j\}_{j\geqslant 1}\in 2^\infty: u_j=s_j, j=1,\cdots,n\}.$$

令$B_n:=\bigcup_{s\in T(n)}f^{-1}(C_s)\times C_s$. 显然，每个$C_s$和$f^{-1}(C_s)$都是博雷尔集，且$T(n)$是有限的，所以$B_n$是一个博雷尔集. 令$G$是所有$B_n(n=1, 2, \cdots)$的交，那么$G$也是博雷尔集. 对每个$y\in 2^\infty$，$n$，$y\in C_s$(唯一的$s=s(n, y)\in T(n)$). 为了证明$G$是$f$的图形，我们有对所有的$n$，$(x, y)\in B_n$当且仅当对所有的$n$，$f(x)\in C_{s(n,y)}$，但是这意味着$f(x)=y$，这就证明了引理13.2.2，从而证明了定理13.2.1. □

或者是空集或者是满足定理13.2.1中等价条件的集合A称为解析集(analytic set). 显然，波兰

空间中的任意博雷尔集都是解析集. 非博雷尔解析集的例子也是非平凡的. 在证明这种集合存在之前，我们能够注意到直接像并不保持某些性质. 如果f是连续函数，U是开集，那么$f[U]$不一定是开的：令$f(x):=x^2$，$U:=(-1,1)$. 如果K是紧的，那么$f[K]$是紧的，但是如果F是闭的，那么$f[F]$不一定是闭的：在$\mathbb{R}^2$中令$F:=\{(x,y):xy=1\}$，f是投影$f(x,y):=x$.

494

非博雷尔解析集的构造是以一些称为通用集的集合为基础的. 令X，Y是集合，S是$X\times Y$的子集，那么对每个$y\in Y$，我们有S的截面(section)，截面是由$\{x\in X:(x,y)\in S\}$定义的X的子集. 令$\mathcal{C}$是X子集的集族，那么S称为$\mathcal{C}$的一个通用集(universal set)，当且仅当$\mathcal{C}$是S的所有截面的集合. 如果$\mathcal{C}$是一个可数集族$\{C_n\}_{n\geqslant 0}$，那么在$X\times\mathbb{N}$中一个简单的通用集是$C:=\{(x,n):x\in C_n,n\in\mathbb{N}\}$. 如果在$X$中每一个集合$C_n$是开的，那么$C$在$X\times\mathbb{N}$中是开集，其中$\mathbb{N}$具有离散拓扑. 但是如果$\mathcal{C}$是不可数集族，一个不可数的离散空间$Y$就不是一个可分度量空间. 由于$\mathbb{N}^\infty$是波兰空间，在定义一个通用开集时，$\mathbb{N}^\infty$通常方便地作为商空间，但是作为离散空间的积，它具有某些"离散"的性质. 例如，$\mathbb{N}^\infty$是完全不连通的：对$\mathbb{N}^\infty$中任意两个不同的点m和n，空间$\mathbb{N}^\infty$可以被记为两个不相交开集U和V的并，其中$m\in U$，$n\in V$. 对于$\mathbb{N}^\infty$还有如下一个通用集.

13.2.3 命题 *对任意的第二可数拓扑空间X，在$X\times\mathbb{N}^\infty$中存在一个开集U(X中所有开集的集族)和一个闭集F，U对于拓扑X来说是一个通用集F对X中所有闭集的集族是一个通集.*

证明 令$\{U_n\}_{n\geqslant 1}$是X的拓扑的可数基，$U_0=\varnothing$. 对所有的n和k，令U是$U_n\times\{\{n_j\}_{j\geqslant 0}:n_k=n\}$的并. 这些集合中的每一个都是开的，所以它们的并$U$也是开集. 令$n(j):=n_j$，$j=1,2,\cdots$对$\mathbb{N}^\infty$中一个给定的点$\{n_j\}$，$U$的对应截面是所有$U_{n(j)}$的并，对给定的序列$n(j)$，其中$n(j)$通过基的定义给出了$X$中所有的开集. 令$F$是$U$的补集，那么就给出了$X\times\mathbb{N}^\infty$上的一个闭集，且它对$X$中所有闭集是一个. □

存在波兰空间X，对X，在$X\times\mathbb{N}^\infty$中对X中所有博雷尔集的集族，不存在通用的博雷尔集(参见本节习题8). 尽管解析集的类比博雷尔集的类要大得多，但是结果表明可以定义通用解析集(它们自己是解析集，但不是博雷尔集).

13.2.4 定理 *对任意的波兰空间X，在$X\times\mathbb{N}^\infty$上存在一个解析集A，对X上所有解析集的集族是通用的.*

495

证明 令F是$(X\times\mathbb{N}^\infty)\times\mathbb{N}^\infty$上的闭集，由命题13.2.3知，对波兰空间$X\times\mathbb{N}^\infty$中所有闭集的集族，$F$是通用的. 令$f(x,\{m_j\},\{n_k\}):=(x,\{n_k\})$是从$(X\times\mathbb{N}^\infty)\times\mathbb{N}^\infty$映射到$X\times\mathbb{N}^\infty$的. 令$A:=f[F]$，那么由定义(由定理13.2.1)知，$A$是解析集. 对每个$n:=\{n_k\}_{k\geqslant 1}\in\mathbb{N}^\infty$，截面$\{x\in X:(x,n)\in A\}$等于

$$\pi[\{(x,m)\in X\times\mathbb{N}^\infty:(x,m,n)\in F\}],$$

其中π是从$X\times\mathbb{N}^\infty$映上到X的自然投影$\pi(x,m):=x$. 对所有闭集$H\subset X\times\mathbb{N}^\infty$，$X$中所有这种截面的集合等于所有$\pi[H]$的集合，因为$F$是一个通用闭集. 现在，任意从$\mathbb{N}^\infty$映射到$X$的连续函数$g$的图形是一个闭集$H$，$H$到$X$上的投影是$g$的值域. 所以由定理13.2.1知，$A$是一个通用解析集. □

现在可以证明非博雷尔解析集是存在的. 注意到，不像勒贝格不可测集存在性的证明那样，下面的证明不会用到选择公理：它给出了一个特殊的也许有些复杂的解析非博雷尔集的例子，并且这个集合比博雷尔集更复杂.

13.2.5 命题 *在任意不可数的波兰空间X中，存在一个解析集A，使得*

(a)A的补集不是解析集，并且

(b)A 不是博雷尔集.

证明 因为(b)很容易由(a)得到，所以只需证明(a). 根据博雷尔同构定理(13.1.1)，可以假设 $X=\mathbb{N}^\infty$. 由定理13.2.4知，在$\mathbb{N}^\infty\times\mathbb{N}^\infty$中取一个解析集 C，对于 X 中的解析集它是通用的. 令 D 是$\mathbb{N}^\infty\times\mathbb{N}^\infty$中的对角元素$\{(n, n): n\in\mathbb{N}^\infty\}$. 我们会看到 $C\cap D$ 是解析的. 由定理13.2.1，令 f 是从$\mathbb{N}^\infty$映上到 C 的函数，那么 $f^{-1}(C\cap D)=f^{-1}(D)$ 是闭的，因此是波兰的，所以根据定理13.2.1知，$C\cap D=f[f^{-1}(C\cap D)]$是解析的. 令 $A:=\{n\in\mathbb{N}^\infty: (n, n)\in C\cap D\}$，则$A=\pi[C\cap D]$，其中 $\pi(m, n):=m$ 是自然投影，所以根据连续函数的复合和定理13.2.1知，A 是 X 中的解析集. 假设它的补集也是解析集，那么它等于一个通用解析集 C 的某个截面，并且对某个 $m\in\mathbb{N}^\infty$，$n\notin A$ 当且仅当$(n, m)\in C$. 当 $m=n$ 时，矛盾. □

尽管解析集不需要是博雷尔集，结果表明对定义在博雷尔集上的任意概率测度的完备化，它们总是可测的. 换句话说，解析集普遍是可测的，正如在11.5节定义的一样. 看起来好像解析非博雷尔集是普遍可测非博雷尔解析集中是最容易得到的例子. 当然，从"普遍"可测的定义可以看出，所有博雷尔集都是普遍可测的.

496

13.2.6 定理 *在任意波兰空间 X 中，任意解析集都是普遍可测的.*

证明 令 A 是解析集，则由定理13.2.1知，$A=f[\mathbb{N}^\infty]$，其中 f 是连续的. 令 μ 是 X 的博雷尔 σ-代数上的任意概率测度. 对任意的正整数 k 和 M，令

$$N(k,M):=\{\{n_j\}_{j\geq 1}\in\mathbb{N}^\infty: n_k\leq M\}.$$

令 $\varepsilon>0$. 因为当 $M\uparrow\infty$ 时，$f[N(1, M)]\uparrow f[\mathbb{N}^\infty]$，由外测度的连续性(定理3.1.11)知，存在一个 M_1，使得

$$\mu^*(f[N(1,M_1)])\geq\mu^*(A)-\varepsilon/2.$$

类似地，重复运用定理3.1.11，对于所有的 k，我们得到 M_k，使得

$$\mu(f[F_k]^-)\geq\mu^*(f[F_k])\geq\mu^*(A)-\sum_{1\leq j\leq k}\varepsilon/2^j>\mu^*(A)-\varepsilon,$$

其中 $F_k:=\bigcap_{1\leq j\leq k}N(j, M_j)$. 当 $k\to\infty$ 时，$F_k\downarrow C:=\bigcap_{j\geq 1}N(j, M_j)$. 每个 F_j 都是闭的，根据吉洪诺夫定理(2.2.8)知，C 是紧的，因为有限集$\{1, \cdots, M_j\}$是紧的. 为了证明 $f[F_k]^-\downarrow f[C]$，将用到定理2.2.12. 对任意开集 $U\supset C$，因为 C 是紧的，$C\subset V\subset U$，其中 V 是积拓扑的定义中积拓扑基集合的有限并(2.2节). 在基中每一个集合仅仅依赖于有限多个坐标. 令 J 是有限子覆盖集合定义中任意坐标的最大指标. 因为 F_J 的点的前 J 个坐标恰好是C中的点，所以有 $F_J\subset U$. 因此，可以运用定理2.2.12，且$\mu(f[C])\geq\mu^*(A)-\varepsilon$. 由于$f[C]$是紧的，取 $\varepsilon=1/n$，$n=1, 2, \cdots$我们得到了紧集的一个可数并，它是一个博雷尔集 B，满足 $B\subset A$，且$\mu(B)=\mu^*(A)$. 因此$\mu^*(A\setminus B)=0$且 A 对 μ 是可测的(根据命题3.3.2). □

如果在集合 C 中对每个 y，$A(y)$是非空的，选择公理指出对所有的 y，在 C 上存在一个函数 g，满足 $g(y)\in A(y)$. 如果每个 $A(y)$是可测集，并且在可测方式上集合 $A(y)$依赖于 y，那么可以取 g 是可测的吗？这里有一个答案，有时称为"截面"定理. 在 $A(\cdot)$上的可测性假设为：在一个积空间中，$A:=\{(x, y): x\in A(y)\}$是一个可测集. 选择公理的这种可测形式将不会依赖于通常的选择公理.

497

13.2.7 定理 *令 X 和 Y 是波兰空间，A 是 $X\times Y$ 的一个解析集子集，C 是 A 映射到 Y 的投影，$C:=\{y\in Y$: 对某个 $x\in X$，$(x, y)\in A\}$，那么存在一个从解析集 C 映射到 X 的函数 g，使得对所有的 $y\in C$，都有*

$(g(y), y) \in A$，且使得从 Y 的普通可测集的 σ-代数映射到 X 的博雷尔集的函数 g 是可测的.

注：在应用中，通常 A 是 $X \times Y$ 上的博雷尔集，那么 C 通常而非必要是 Y 的博雷尔集.

选择公理与良序原理是等价的(定理 1.5.1). 在$[0, \infty]$上，每一个非空闭集都有一个最小元. 在$\mathbb{N}^\infty$中，良序集的这种比较弱的形式将会对"可测选择"有帮助.

证明 $\mathbb{N}^\infty$ 上的字典序定义为 $\{m_j\}_{j\geq 1} < \{n_j\}_{n\geq 1}$，当且仅当对某个 i，且对所有的 $j < i$，$m_j = n_j$，$m_i < n_i$. 那么，我们有下面的引理.

13.2.8 引理 $\mathbb{N}^\infty$ 的任意非空闭子集 F 有最小元.

证明 令 n_1 是 F 中任意点的第一个最小的坐标，并且给定 $n_1, \cdots, n_{k-1}$，令 n_k 是 F 中任意点的具有前 $k-1$ 个坐标 $n_1, \cdots, n_{k-1}$ 的第 k 个最小的坐标. 对每个 k，存在 F 的一个点 $m^{(k)}$，它的前 k 个坐标等于 $n_1, \cdots, n_k$，并且这些 $m^{(k)}$ 收敛到$\{n_j\}_{j\geq 1}$，因此，它们一定在 F 中且是最小元. □

现在继续定理 13.2.7 的证明，由定理 13.2.1，令 f 是从$\mathbb{N}^\infty$映上到 A 的连续函数. $\pi_1(x, y) := x$，$\pi_2(x, y) := y$，那么 $\gamma := \pi_2 \circ f$ 是从$\mathbb{N}^\infty$映上到 C 的连续函数. 所以根据定理 13.2.1 可知，C 是解析的，并且根据定理 13.2.6 知，C 是普遍可测的. 对每个 $y \in C$，$\gamma^{-1}(\{y\})$ 在$\mathbb{N}^\infty$中是闭的. 令 $h(y)$ 是它的最小元，由引理 13.2.8 知，如果 h 在 Y 中对于普遍可测的 σ-代数是可测的(在其值域是博雷尔 σ-代数)，那么 $g := \pi_1 \circ f \circ h$ 也是可测的，从而具有期望的性质. 所以只需要证明 h 的可测性.

对某个满足 $m_1 = n$ 的 $m = \{m_j\}_{j\geq 1} \in \mathbb{N}^\infty$ 和对所有 $y \in C$，使得 $\gamma(m) = y$ 的集合 $C(n)$，显然 $C(n)$ 是一个解析集. 令 $h_1(y)$ 是使得 $y \in C(n)$ 的最小的 n. 那么 $h_1(y) = n$ 当且仅当 $y \in C(n) \bigcup_{j<k} C(j)$，所以 h_1 对由解析集生成的 σ-代数 $\mathcal{S}(\mathcal{A})$ 是可测的. 递归地，给定 $h_1, \cdots h_k$，对某些满足对 $j \leq k$ 有 $m_j = h_j(y)$ 和 $m_{k+1} = n$ 的 m，令 $h_{k+1}(y)$ 是使得 $\gamma(m) = y$ 的最小的 n. 那么每一个 h_k 都是可测的，并且 $y \mapsto \{h_k(y)\}_{k\geq 1}$ 也是可测的，由此完成定理 13.2.7 的证明. □

498

习题

1. 对每一个 $n = 1, 2, \cdots$ 令 A_n 是波兰空间 S 中的解析集，证明：$\bigcup_n A_n$ 和 $\bigcap_n A_n$ 也是解析集. [提示：参见定理 13.1.5 的证明.]
2. 集合 B 称为 F_σ，如果它是闭集的一个可数并. 令 B 是$\mathbb{R}^k$ 中的 F_σ，f 是从 B 映射到度量空间 S 的一个连续函数，证明：$f[B]$是 S 中的 F_σ.
3. 开集的一个可数交称为 G_δ(集). 给出一个$\mathbb{R}$ 中 G_δ 集 B 且从 B 映射到度量空间$\mathbb{R}$ 使得$f[B]$不是一个 G_δ(集)的连续函数 f 的例子.
4. 对某些豪斯多夫拓扑空间 X 和 Y，令 f 是从 X 映射到 Y 的一个连续函数. 证明：f 的图形在$X \times Y$ 中是闭的且具有积拓扑.
5. 令 f 是从紧豪斯多夫空间 K 映射到另一个同样空间 L 的函数，使得 f 的图形在 $K \times L$ 中是闭的. 证明：f 是连续的. [提示：运用定理 2.2.11.]
6. 给出一个从紧度量空间 K 映射到度量空间 Y 的函数 f 的例子，使得 f 的图形在 $K \times Y$ 中是闭的，但是 f 是不连续的. [提示：K 为单位圆，$Y = [0, 2\pi)$.]
7. 令$(X, \mathcal{B})$是可测空间，证明：对 $X \times X$ 中的积 σ-代数$\mathcal{B} \otimes \mathcal{B}$上的任意集合 A，$\{x: (x, x) \in A\}$ 在$\mathcal{B}$中. [提示：矩形、半环、代数和单调类.]
8. 对任意的可测空间$(X, \mathcal{B})$，证明：在 $X \times X$ 中的积 σ-代数$\mathcal{B} \otimes \mathcal{B}$上，不存在集合 A 对$\mathcal{B}$是通用的. [提示：参见习题 7 和命题 13.2.5 证明的结尾部分]. 注：在定理 13.2.4 中，对有博雷尔 σ-代数的 $X = \mathbb{N}^\infty$，"解析

的"不能被"博雷尔的"代替.

9. 令整数空间$\mathbb{Z}$有离散拓扑，$\mathbb{Z}^{\mathbb{N}}$是所有从$\mathbb{N}$映射到$\mathbb{Z}$的函数的集合，具有积拓扑，那么$\mathbb{Z}^{\mathbb{N}}$是一个波兰空间. 在$\mathbb{Z}^{\mathbb{N}}$上，对$n\in\mathbb{N}$，令$\mathcal{F}_n$是在其上$z=\{z_n\}_{n\geqslant 0}\in\mathbb{Z}^{\mathbb{N}}$的坐标$z_0$，…，$z_n$可测的最小$\sigma$-代数. 停时是一个从$\mathbb{Z}^{\mathbb{N}}$映射到$\mathbb{N}\cup\{+\infty\}$的函数$\tau$，使得对所有的$n$，$\{\tau\leqslant n\}\in\mathcal{F}_n$. 在$\mathbb{N}\cup\{+\infty\}$上，构造一个拓扑使得对每个$n<\infty$，$\{n\}$是开的，并且$\infty$的每一个邻域都包含至多有限个$n$. 在从$\mathbb{Z}^{\mathbb{N}}$映射到$\mathbb{N}\cup\{+\infty\}$的所有函数的集合$K$上，用积拓扑. 证明：

499

(a)K是紧的.

(b)所有停时的集合作为K的一个闭子集F也是紧的.

(c)K不是可度量的，但F是. [提示：每一个$t\in F$是唯一地由使得对某个m和所有的$n\geqslant m$，$z_n=z_m$的序列$\{z_n\}$所决定.]

10. (连续性). 令T是使得对某个$f\in\mathbb{Z}^{\mathbb{N}}$，$\tau(f)=+\infty$的所有停时$\tau$的集合. 证明：$T$是解析集. 注：$T$不是博雷尔集.

注释

13.1 节 本节是基于 Parthasarathy(1967，第 1 章，§2)的研究，他参考了 Kuratowski(1933；1966，p. 489). Kuratowski 反过来给出了几个参考文献. 完备可分度量空间上的一个不可数博雷尔集有一个完全子集(定理 13.1.7)，这个结论已由 Alexandroff(1916)和 Hausdorff(1916)分别独立地证明了. 对于其他的事实，可参考 Lusin(1930，第 2 章，参见 p. 114)和 Kuratowski(1934).

13.2 节 Lebesgue(1905，p. 191—192，195—196)认为他证明了对(x, y)平面中的每一个博雷尔集，在x轴上的投影是直线上的博雷尔集. Suslin(1917)发现了证明中的错误，引进了解析集，并且发现在直线上存在一个解析集，但是它的补集不是解析集也不是博雷尔集. 那时 Suslin 是 N. Lusin 的学生. Siepiński(1950，p. 28)写到："我偶然一次碰到 Michel Suslin 正在与 Lusin 女士讨论他的研究结果，Suslin 告诉 Lusin 他发现了一个非常著名结论中的一个错误……". 从 Yushkevich(1968，p. 575)我们了解到 Suslin 出生于 1894 年 11 月 15 日卒于 1919 年，年仅 25 岁. 他几乎不可能发表 1917 之后所取得的结果，尽管关于线性有序集问题的一个简短陈述(Suslin，1920)仍然以他的名字命名. 例如，参见 Mauldin 和 Ulam(1987，p. 285).

Lusin(1917)在 Suslin 之后的一个笔记中指出在直线上的每一个解析集都是勒贝格可测的，这是波兰空间中的解析集都是普遍可测的一个主要例子.

Paplauscas(1973)关于 Lusin(1883—1950)写到："在 1914—1924 年期间，一个智慧的学者和科学的组织者 Lusin，是莫斯科学校里研究实函数理论的核心……，如杰出的数学家 P. S. Alexandrov、A. Y. Khinchin、D. E. Menshov、M. Y. Suslin、A. N. Kolmogorov、N. K. Bari 和 P. S. Novikov 都是他的学生". Lusin(1930)写了一本关于解析集的书，勒贝格在前言中写到："正是他们发现了这个错误，才取得了更进一步的研究成果".

回想到解析集A集及其补集CA集，如果从$\mathbb{R}^2$上的CA集投影到$\mathbb{R}$上的集合，就可以得到一个集合称为PCA集合，它既不是A集也不是CA集. 交选补集和投影给出"投影"集合的层次，被 Sierpiński(1950)解决了.

Cohn(1980，第 8 章)是本节的某些事实和证明源头，而且进一步介绍了解析集. Dellacherie(1978)给出了在概率理论中有用的(非博雷尔)解析集的例子.

500

参考文献

Alexandroff [Aleksandrov], P. (1916). Sur la puissance des ensembles mesurables B. *Comptes Rendus Acad. Sci. Paris 162*: 323–325.

Cohn, Donald Lee (1980). *Measure Theory*. Birkhäuser, Boston.

Dellacherie, Claude (1978). *Quelques exemples familiers, en probabilités, d'ensembles analytiques, non boréliens, Séminaire de Probabilités, Univ. Strasbourg* (1976–1977); *Lecture Notes in Math.* (Springer) *649*: 746–756.

Hausdorff, Felix (1916). Die Mächtigkeit der Borelschen Mengen. *Math. Ann. 77:* 430–437.

Kuratowski, Kazimierz [Casimir] (1933). *Topologie I.* Monografje Matematyczne, Warsaw and Lwow.

——— (1934). Sur une généralisation de la notion d'homéomorphie. *Fund. Math. 22*: 206–220.

——— (1966). *Topology*, vol. 1, transl. from 2d French ed. (1958) by J. Jaworowski. Academic Press, New York; PWN, Warsaw.

Lebesgue, Henri (1905). Sur les fonctions représentables analytiquement. *Journal de Mathématiques* (Ser. 6) *1*: 139–216.

Lusin [Luzin], Nikolai Nikolaevich (1917). Sur la classification de M. Baire. *Comptes Rendus Acad. Sci. Paris 164*: 91–94.

——— (1930). *Leçons sur les ensembles analytiques et leurs applications*, with a Note by W. Sierpiński and a preface by H. Lebesgue. Gauthier-Villars, Paris.

Mauldin, R. Daniel, and Stanislaw M. Ulam (1987). Mathematical problems and games. *Adv. Appl. Math. 8*: 281–344.

Paplauscas [Paplauskas], A. B. (1973). Luzin, Nikolai Nikolaievich. *Dictionary of Scientific Biography, 8*: 557–559.

Parthasarathy, K. R. [Kalyanapuram Rangachari] (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

Sierpiński, Wacław (1950). *Les ensembles projectifs et analytiques*. In *Mémorial des Sciences Mathématiques*, fasc. 112. Gauthier-Villars, Paris.

Souslin [Suslin], Mikhail Yakovlevich (1917). Sur une définition des ensembles mesurables B sans nombres transfinis. *Comptes Rendus Acad. Sci. Paris 164*: 88–91.

——— (1920). [Problem]. *Fund. Math. 1*: 223.

Yushkevich, A. P. (1968). *Istoriya Matematiki v Rossii do 1917 goda* [in Russian]. Nauka, Moscow.

附录 A　公理化集合论

尽管该附录比 1.1 节介绍的更详细，更有条理，但没有表明要给出公理化集合论的一个完整严密的表述(有些书中完整地研究了该主题，其中的一些书列在了参考文献里). 尽管这里将要详细阐述集合论的公理，但是将假设读者已经了解一些定义(如线性序和良序)(见第 1 章).

A.1　数理逻辑

大约公元前 300 年，欧几里得几何学从集合的定义、假定和公理，提出了“一种严格地逻辑推理理论”(Struik，1948，p. 59). 尽管欧几里得没有走过所有的路，但也走了很长一段路到才给出公理化方法的现代思想，当定理的证明被详细写出时，它可以机械而精确地进行检查以确定其是否是一个证明. 从现代的观点来看，也许欧几里得系统里最不严密的逻辑部分就是它的定义——例如，“点是没有扩张的”，“直线有长度没有宽度……”正如 1.1 节所注释的，一个真正准确的数学系统，或者“形式系统”是从一些基本的不定项开始的. 从而其他的项就可以依据基本项定义.

给出现代数学基础、被广泛接受的形式系统是以命题演算和一阶谓词逻辑为基础的. 这里只简单介绍了这些主题. 更详细的介绍参见 Kleene(1967).

在命题演算中，陈述与其他陈述是用命题联结词结合起来的. 两个基本的命题联结词是“与”和“非”. 其他常用的联结词可用从这两个联结词定义. 令 P 和 Q 为两个陈述，如“$x=y$”，“$u\neq v$”. 那么 P 或者 Q 定义为非(非 P 且非 Q).

最后一个定义表明给出精确的数学表达式有一些困难. 在数理逻辑中，使用如下的引证符号：用引证符号开始和结束的任意表达式是没有这些引证符号的表达式的一个名字. 因此，“A”表示字母 A，“$A(B)$”表示四个字母的字符串：A，左圆括号，B，右圆括号. (当以这种方式使用通常的双引证符号(”)时，单引号(‘，’)可用于其他引证符号的使用.)如果“或”的定义已经用引证符号表示，比如说“P 或 Q”意味着“非(非 P 且非 Q)”，这已经不需指定，仅应用于引证符号内部的表达式，包含真实的字母“P”和“Q”. 进一步地，这些字母甚至不是陈述. 这意味着 P 和 Q 应该是延展陈述的变量，所以“P”和“Q”中的每一个都可用一个陈述替换. 为了对该定义给出一个方便且精确的形式，逻辑学家 W. V. Quine 发明了一种引证符号变量⌜和⌝，称为‘角标’. 例如，⌜P 和 Q⌝表示⌜非(非 P 且非 Q)⌝. 这里，在角标内部的两个表达式，当 P 用一个陈述代替，Q 用另一个陈述代替时，结果被认为是一个定义. 只有在进行了这种替换后，才可真正正确地使用通常的引证符号，例如，“$x=y$ 或者 $u=v$”表示“非($x\neq y$ 且 $u\neq v$)”.

在形式系统里的陈述和关于系统的陈述之间需要加以区别. 关于系统的陈述可能包含‘语法变量’，比如在系统里关于陈述或者其他表达式变化的“P”或者“Q”. Quine 角标是一种由简单变量形成复语法变量的方法.

现在来定义其他重要的联结词，⌜$P\Rightarrow Q$⌝读作⌜P 蕴涵 Q⌝或者如果⌜P，则 Q⌝，它是通过⌜非(P 且非 Q)⌝定义的. 同样，⌜$P\Leftrightarrow Q$⌝是通过⌜$(P\Rightarrow Q)$ 和 $(Q\Rightarrow P)$⌝定义的.

命题演算有其自己的公理和推理法则，这并入其他更加复杂的形式系统. 例如，推理的一个法则是‘假言推理’：

由「P」和「$P\Rightarrow Q$」,可推出「Q」.

这里似乎没有必要阐明其他的命题公理或者法则，因为它们也对应于的数学推理模式.

实现公理化集合论的下一个主要阶段是一阶谓词逻辑或者谓词演算. 这里有两个称为量词的逻辑符号：“∀”读作‘对任意的’，“∃”读作‘存在’. 使用量词“∀”和“∃”、变量(比如“x”和“y”)、命题联结词“非”和“与”以及(由如上这两个可定义的)其他词、符号“=”、“≠”和圆括号，我们可以构建‘合式公式’或者简称‘wffs’. 关于什么是合式公式，有一个准确的语法，或者规则集合. 在其他的情况中，左、右圆括号必须像通常一样恰当地匹配. 例如，

$$“(\forall x)(\exists y)((y=x)\text{ 与 }(y\neq z))”$$

是一个合式公式. 在该公式中，当一个变量(比如“x”或者“y”)由一个量词控制时，称为‘约束的’. 否则，像此公式中的“z”，称为‘自由的’.

再次重申，命题演算有其自己的公理和推理法则，这些将在公理化集合论的完全表示中包括. 在这些表达式中间存在“替换法则”，也就是一个表达式可以代替另一个表达式，例如，在哪些约束变量上可以保持顺序，如何保持这些顺序，以及哪些变量是自由的. 这些法则在这个领域(Kleene，1967，p. 107；Church，1956，p. 289—290)中引起一些技术难题，但似乎不能反应在分析或者概率领域内研究人员需要注意的问题.

A.2 集合论的公理

在 Ernst Zermelo 于 1908 年提出了公理化系统并且 Abraham Fraenkel 改进了该系统之后，最广泛使用的集合论公理系统称为 ZF，或者 Zermelo-Fraenkel. 在 ZF 中，仅有的对象是集合. 它们之间只有一个基本关系，记作“∈”. 例如，如果 x 和 y 是集合，则“$x\in y$”读作‘x 是 y 的一个元素’或者‘x 属于 y’. 关系的否定记作“$x\notin y$”，读作‘x 不是 y 的一个元素’. 在公理化集合论的一个形式系统(比如 ZF)中，‘集合’和‘∈’开始作为未定义的概念，通过公理得到内涵. 为了给‘集合变量’一个准确的定义——即一个在集合上变化的变量——可以仅使用表达式“x”，“x'”，“x''”，…. 但是这里，少些形式，也允许其他变量，如“y”和“X”.

在 ZF 中,「$\forall x$」意思是「对所有的集合 x」,「$\exists y$」意思是「存在一个集合 y[使得……]」.

支配合式公式的语法包括“∈”的使用法则. 每一个“∈”必须立即或者被一个集合变量或者被一些其他的表达式置前和跟随，而表达式允许起相同的作用. 更详细的内容参见 Cohen(1966，p. 6—7). 例如，

$$“(\forall x)(\exists y)((x\in y)\Rightarrow(x\in z))”$$

是一个合式公式，而

$$“(\forall x)(\exists z)\forall y(x\in yz)\exists\in”$$

由于一些原因不是一个合式公式.

对于 ZF，公理由一种表示到另一种表示发生一些变化. 下面是一种斜述，由 9 个公理组成(不包括选择公理).

ZF1(外延性) $\forall y\forall z((y=z)$ 当且仅当 $\forall x(x\in y$ 当且仅当 $x\in z))$.

ZF1 表明两个集合是相等的当且仅当它们有相同的元素.

ZF2(空集) $(\exists y)(\forall x)x\notin y$.

ZF2 说明某个集合 y 没有元素. 由外延性知，恰好存在这样一个集合. 该空集记为∅.

ZF3(无序对)　$\forall x\forall y\exists z\forall u((u\in z)$当且仅当$((u=x)$或者$(u=y)))$.

由外延性知，在 ZF3 中仅有元素 x 和 y 的集合 z 是唯一的. z 记为$\{x, y\}$. 注意到对任意的 x 和 y, $\{x, y\}=\{y, x\}$. 这里$\{x, y\}$称为具有元素 x 和 y 的一个无序对. 置 $x=y$，令$\{x\}$表示$\{x, x\}$. 那么$\{x\}$是恰有元素 x 的一个集合. 这里$\{x\}$读作'单元素集 x'.

下面的不仅是一个公理，而且是一个公理模式：它描述了(无限多个)合式公式是如何形成的，它们都将被看作公理.

ZF4(选择)　对每一个仅包含一个自由变量"x"的合式公式⌈P⌉，且不包含变量"y"或"z"，下面的是一个公理:⌈$\forall y\exists z\forall x(x\in z$当且仅当$(x\in y$且$P))$⌉.

注意到角标(不是引号)是需要的，因为合式公式 P 是变化的.（在"x"，"y"和"z"周围引号的位置用角标替换将允许更多的公式作为公理，但将给出定义无附加集合的等价公理.）由外延性知，在 ZF4 中集合 z 是唯一的. 它将记为⌈$\{x\in y: P\}$⌉，读作⌈y 中所有 x 的集合满足 P⌉. ZF4 通常称为 Aussonderung 公理，来自意思为'选择'的德语词：从集合 y，我们可以选择满足 P 的所有元素形成一个集合. 但在 ZF4 中，对某个集合的子集的限制是需要的以避免矛盾. 没有这一点，如果试图定义，例如，$r=\{x: x\notin x\}$，则 $r\in r$ 意味着 $r\notin r$ 且反之亦然(罗素悖论).

下一个公理允许集合的任意集 x 的并集 u 的形成，其中集合的集族在集合理论中仅是一个集合.

ZF5(并集)　$\forall x\exists u\forall y((y\in u)$当且仅当$(\exists z)((y\in z)$且$(z\in x)))$.

由外延性知，一旦给定 x，集合 u 是唯一确定的. u 称为 x 的并集，且记作 $\cup x$. 这个简单的符号在集合论中经常使用，但在实分析或其他领域中不太常用. 另一个可能的符号是 $\bigcup_{y\in x} y$. 我们也可以定义两个集合的并：令 $x\cup y=\cup\{x, y\}=\{z: z\in x$ 或者 $z\in y\}$.（第一项 $x\cup y$ 是被定义项，根据 ZF3 和 ZF5，中间项 $\cup\{x, y\}$ 是明确定义的，尽管在符号上类似于 ZF4 中的，但是形式上最后一个描述不被 ZF4 所允许.）

从无序对和并集公理出发，对任意的 x_1, x_2, x_3 和 x_4，可以得到拥有给定四个元素的一个集合 $\{x_1, x_2, x_3, x_4\}$. 迭代此过程，可得到包含集合的任意给定有限集合列的集合.

定义　对任意包含一个自由变量⌈x⌉的合式公式⌈Q⌉,⌈$(\forall x\in y)Q$⌉意味着⌈$\forall x((x\in y)\Rightarrow Q)$⌉. 那么⌈$y\subset z$⌉，读作'$y$ 包含在 z 中'，意味着⌈$(\forall x\in y)x\in z$⌉.

ZF6(幂集)　对任意集合 x，存在一个集合 z，使得 $y\in z$ 当且仅当 $y\subset x$.

在 ZF6 中，集合 z 是 x 的所有子集的集合，通常是唯一的，有时记作 2^x.（如果 x 中有 n 个元素，则 2^x 中有 2^n 个.）对任意两个集合 x 和 y，它们的交集定义为 $x\cap y=\{u\in x: u\in y\}$，根据 ZF4 知，$x\cap y$ 是一个集合. 那么称 x 和 y 是不相交的，当且仅当 $x\cap y=\varnothing$. 如果 x 是任意非空集，则它的交集定义为

$$\cap x=\bigcap_{v\in x} v=\{u: u\in v, \forall v\in x\}.$$

根据 ZF4(选择)知，这也是一个集合，例如，我们可以取$\{u\in y: \cdots\}$，其中 $y=\cup x$. 差 $x\setminus y$ 定义为 $x\setminus y=\{u\in x: u\notin y\}$.

至此，我们不能用公理证明存在一个无限集. 下面的公理证明了存在一个无限集.

ZF7(无限集)　$\exists M(\varnothing\in M$ 且$(\forall x\in M)x\cup\{x\}\in M)$.

正如在 ZF7 中，集合 M 包含这样的元素$\varnothing$, $\{\varnothing\}$, $\{\varnothing, \{\varnothing\}\}\cdots$. 在集合论中构建非负整数

的标准方法是分别定义等于0，1，2，…的如上集合. 所有非负整数的集合$\mathbb{N}$将定义为满足 ZF7 的条件的最小集合 M，下面的定理证明了这样的集合是存在的.

A.2.1 定理 *存在满足 ZF7 中关于 M 的条件的唯一集合$\mathbb{N}$，使得对任意其他这样的集合 M，$\mathbb{N}\subset M$.*

证明 令 M 是任意满足 ZF7 的集合，$H(M)$是也满足条件的 M 的所有子集的集合. 那么根据选择(ZF4)知，$H(M)$的确是一个集合，因为它是 2^M 的子集. 令 $N(M)=\cap H(M)$，则 $N(M)\subset M$. 如果 $J\in H(M)$且 $K\in H(M)$，那么 $J\cap K\in H(M)$. 从而有 $N(J)=N(M)$. 同理有 $N(M)\in H(M)$. 现在令 M'是任意满足 ZF7 的其他集合，则 $M\cap M'$也满足 ZF7. 因此 $N(M)=N(M\cap M')=N(M')$. 所以 $N(M)$不依赖于 M，且可以置$\mathbb{N}:=N(M)$. 那么$\mathbb{N}\subset M$，所以定理中描述的集合是存在的. 令 K 为任意这样的集合，则 $K\subset\mathbb{N}$ 且$\mathbb{N}\subset K$，所以$\mathbb{N}=K$，证毕. □

定义 一个**有序对**是对某个集合 x 和 y 形如$\{\{x\},\{x,y\}\}$的集合. 对于任意的有序对 Q，它的**第一个元素**(first member)是使得$\{x\}\in Q$ 的唯一的 x. 它的**第二个元素**(second member)是使得 $Q=\{\{s\},\{x,y\}\}$的唯一的 y，其中 x 是第一个元素. 令$\langle x,y\rangle$表示$\{\{x\},\{x,y\}\}$.

上面传统的定义仅是解码期望对的一种方法. 顺便注意到$\{\{x\},\{x,y\}\}=\{\{y,x\},\{x\}\}$，等等.

定义 对于任意两个集合 X 和 Y，它们的笛卡儿积 $X\times Y$ 定义为集合$\{\langle x,y\rangle: x\in X, y\in Y\}$.

根据选择公理 ZF4，$X\times Y$ 是一个集合，作为 2^U 的子集，其中 $U=2^{X\cup Y}$. 关系是有序对的任意集合. 对任意关系 E，逆关系定义为 $E^{-1}:=\{\langle y,x\rangle:\langle x,y\rangle\in E\}$. 根据 ZF4，这也是一个集合，因为它是 2^U 的一个子集.

定义 一个**函数**是一个集合 f，它的所有元素是有序对，使得对任意的 x，如果$\langle x,y\rangle\in f$ 且$\langle x,z\rangle\in f$，那么 $y=z$. 对任意的函数 f，用 $f(x)=y$ 来表示$\langle x,y\rangle\in f$. f 的**定义域**定义为所有 x 的集合，在其上 $f(x)$是有定义的，换句话说，$\mathrm{dom}\, f=\{x:\exists y\langle x,y\rangle\in f\}$. 这里运用 ZF4，因为可以用$\{x\in\cup\cup f:\cdots\}$代替$\{x:\cdots\}$. f 的**值域**是其值的集合：$\mathrm{ran} f:=\{y:$ 对某个 x，$f(x)=y\}$，它也是$\cup\cup f$ 的一个子集. 这里"$:=$"总是表示"通过定义相等".

假设有一个含两个自由变量 x 和 y 的合式公式，记为⌈$E(x,y)$⌉，且对集合 t 中的每个 x，$E(x,y)$对 y 的最多一个值成立. 那么据选择公理，$\{x\in t:\exists yE(x,y)\}$是一个集合. 为了定义一个函数 f，使得$(f(x)=y)$等价于$(E(x,y)$且 $x\in t)$，这只需$\{y:(\exists x\in t)E(x,y)\}$是一个集合就足够了.（注意：$(\exists x\in t)\cdots$意味着$(\exists x)(x\in t$ 且$\cdots)$.）因为否则这不是显然的，下一步将被假设，其中这个函数可能也依赖于其他变量. 正像 ZF4，这将不仅是一个公理，而且是一个公理模式. 这里 $E(u,v)$表示通过 $E(x,y)$将 u 代替 x、v 代替 y 的结果. 例如，如果 $E(x,y)$是"$(\forall t)(t\in x$ 或者 $t\notin y)$"，则 $E(u,v)$是"$(\forall t)(t\in u$ 或者 $t\notin v)$"，对包含更多变量的合式公式这也同样成立.

ZF8(替换或代换) 令⌈$E(x,y,x(1),\cdots,x(k))$⌉是一个含有自由变量"x"，"y"，"$x(1)$"，"$x(2)$"，…，"$x(k)$"的合式公式，其中 $k=0,1,\cdots$，且没有其他的自由变量，并且不包含约束变量"t"，"u"，"v"，"w"或"z"，那么以下形式的每一种陈述是一个公理：

$$\forall x(1)\cdots\forall x(k)(\forall x\forall y\forall z((E(x,y,x(1),\cdots,x(k)),$$
$$E(x,z,x(1),\cdots,x(k))\Rightarrow y=z)))\Rightarrow\forall t\exists w\forall v(v\in w \text{ 当且仅当 }$$
$$(\exists u)(u\in t \text{ 且 } E(u,v,x(1),\cdots,x(k)))).$$

在一种意义下，y 是 $k+1$ 个变量 x，$x(1)$，…，$x(k)$的函数．在另一种意义下，y 是 x 的函数，其中 $x(1)$，…，$x(k)$是参数．例如，关系 E(一旦定义了代数运算)可能是 $y=ax^2+bx+c$，其中 $k=3$，$a=x_1$，$b=x_2$ 和 $c=x_3$．公理 ZF8 在构建序数和某个大集合时是有用的(参见 A.3 节)．例如，如果从 ZF7 取一个无限集 M，则可以形成 2^M，然后形成 2^{2^M}，等等．假设想要定义一个集合 X，使得$\mathbb{N}\in X$且对所有的 $A\in X$，有 $2^A\in X$．我们能够写出的集合的一个‘可数并’也应该是一个集合，这看起来是自然的，但它并不遵循 ZF1～ZF7 所述．如果包含 ZF8 则是遵循的．这种大集合在数学中几乎是不需要的，且在本书中也没有使用．在寻找一个很一般的断言的证明而未成功时，我们可以通过在每个很大的实例集合中找一个反例，并由此可以得到一些启发．

又一个公理有助于简化和明确集合论，它是通过防止集合成为相互的元素或者从有无限元素链得到的：例如，$\cdots\in x(3)\in x(2)\in x(1)$．

ZF9(正则性)　对任意 $x\neq\varnothing$ 的集合，$(\exists y\in x)x\cap y=\varnothing$．

A.2.2 定理　对所有的 x 和 y，$x\notin x$ 且如果 $x\in y$，则 $y\notin x$．

证明　根据 ZF9，$x\cap\{x\}=\varnothing$．如果 $x\in x$，则 $x\in x\cap\{x\}$，矛盾．所以 $x\notin x$．(对 $x=y$，这个结论也可以从定理的第二部分得到．)如果 $x\in y\in x$，那么由 ZF9 可知，在$\{x, y\}$中有一个元素与$\{x, y\}$不相交，由对称性可知，此元素为 x．因此，$y\in x\cap\{x, y\}$，矛盾．□

通常(但不总是)假设剩余公理是选择公理和推广的连续统假设，这在 1.5 节中解决了．

习题

1. 对给定的有序对定义，求(a)$\bigcup\langle x, y\rangle$；(b)$\bigcup\bigcup\langle x, x\rangle$．
2. 如果 $x\in y\in z\in w$，证明 $w\notin x$．
3. 证明：在$\mathbb{N}$上，不存在函数f，使得对所有的 $n\in\mathbb{N}$，$f(n+1)\in f(n)$．[提示：把 ZF9 应用到f的值域．]
4. 证明：如果 A 和 B 是集合，那么 $A\times B$ 也是集合．
5. 证明：如果 E 是一个关系，那么 E^{-1}也是关系．
6. 利用 ZF8 详细证明：存在一个满足$\mathbb{N}\in X$的集合 X 且使得对每个 $B\in X$，都有 $2^B\in X$．

A.3　序数和势

两个偏序集$(X, \leqslant)$和$(Y, \leqslant)$称为*序同构的*(order-isomorphic)，当且仅当存在一个从 X 映射到 Y 的一一对应函数f，使得对任意的 $x\in X$ 和 $t\in X$，$x\leqslant t$ 当且仅当$f(x)\leqslant f(t)$．函数f称为一个*序同构*．序同构是一个等价关系(在任意偏序集上)．选择良序集的方法称为序数，使得每个良序集都序同构于一个良序集且仅一个序数．下面给出一个简单的定义：集合 x 称为是全的(full)当且仅当对所有的 $y\in x$，有 $y\subset x$．一个被$\in$线性地规定的序数是一个全集，所以在 X 上$\{\langle y, z\rangle\in X\times X: y\in z\}$是一个线性序．(根据定理 A.2.2，这个序一定是严格的．)对于序数，$x<y$ 意味着 $x\in y$．

显然，空集$\varnothing$是一个序数($\in$完全地支持了全集和线性序)．容易验证，如果 x 是一个序数，那么 $x\cup\{x\}$也是序数．令$\{a, b, c\}$是一个仅含元素 a，b 和 c 的集合，对于任意包含有限个元素的集合也可以类似处理．(根据无序对和并公理，这样的集合存在．)因此，下面的都是序数：$\varnothing$，$\{\varnothing\}$，$\{\varnothing, \{\varnothing\}\}$，$\{\varnothing, \{\varnothing\}\{\varnothing, \{\varnothing\}\}\}$，…从序数的观点看，这些集合称为 0，1，2，3，…所以每一个非负整数 n 都与一个有 n 个元素的序数对应．为了处理无限序数，需要给出更多的结论．

A.3.1 定理　每个序数 z 在$\in$下都是良序的．

证明 由定义知，z 在 $\in$ 下是线性序的. 令 $A\subset z$，$A\neq\varnothing$，由 ZF9，取 $x\in A$，使得 $x\cap A=\varnothing$. 那么对所有的 $y\in A$，有 $y\notin x$，所以 $x\in y$ 或 $x=y$. 故 x 是 A 的最小元素. □

A.3.2 命题 *如果 $u\in v$ 且 v 是一个序数，那么 u 也是一个序数.*

证明 因为 v 是全的，$u\subset v$，所以 u 在 $\in$ 下是线性序的. 如果 $x\in u$，那么 $x\in v$，所以 $x\subset v$. 如果 $x\not\subset u$(x 不包含在 u 内)，取 $y\in x\setminus u$，那么 $y\in v$ 且 $y\notin u$. 因为 v 在 $\in$ 下是线性序的，所以 $y=u$ 或 $u\in y$. 如果 $u=y$，那么由 $y\in x\in u$ 知，这与定理 A.2.2 矛盾. 如果 $u\in y$，那么 $u\in y\in x\in u$，与应用于$\{x,y,u\}$的正则性公理(ZF9)矛盾. 因此 . u 是全的. □

对任意偏序集$(X,<)$，一个初始段是满足只要 $x<y\in A$ 就有 $x\in A$ 的集合 $A\subset X$.

A.3.3 命题 *对任意的序数 x 和 y，$x\cap y$ 都是 x 的一个初始段.*

证明 如果 $u<v\in x\cap y$，那么 $u\in v\in x$ 且 $v\in y$. 因为 x 和 y 是全的，$v\subset x\cap y$，所以$u\in x\cap y$. □

A.3.4 命题 *任意序数集在包含关系下都是线性序的.*

证明 如果 u 和 v 是序数，$u\not\subset v$ 且 $v\not\subset u$，设 s 和 t 分别是 $u\setminus v$ 和 $v\setminus u$ 的最小元. 如果 $y\in s$，那么由 u 是全的知，$y\in u$ 且 $y<s$，因此 $y\notin u\setminus v$ 且 $y\in u\cap v$. 因此 $s\subset u\cap v$. 类似地，有 $t\subset u\cap v$. 反之，由命题 A.3.3 知，如果 $x\in u\cap v$，那么 $x<s$. 换句话说，$x\in s$，所以 $u\cap v\subset s$. 故 $s=u\cap v=t$，矛盾. □

A.3.5 命题 *任意序数集在 $\in$ 关系下都是线性序的.*

证明 假设 u 和 v 是满足 $u\subset v$ 且 $u\neq v$ 的序数. 由定理 A.3.1 知，令 y 是 $v\setminus u$ 的最小元，那么由命题 A.3.2 和 v 在 $\in$ 关系下是线性序的知，$u\subset y$. 对任意的 $x\in y$，我们有 $x\in v$，由 y 的定义知，$x\in u$. 所以 $u=y\in v$. □

A.3.6 推论 *序数的任意集合 S 在 $\in$ 下都是良序的.*

证明 设 $A\subset S$，$A\neq\varnothing$. 由正则性(ZF9)，取 $x\in A$ 且满足 $x\cap A=\varnothing$，那么由命题 A.3.5 可得，x 是 A 中的最小元. □

A.3.7 定理 *任意序数集 S 的并仍是一个序数.*

证明 回想序数的所有元素仍是序数(命题 A.3.2). 设 U 是并集，且 x，$y\in U$. 设 $x\in u\in S$ 且 $y\in v\in S$. 由命题 A.3.4，我们可取 $u=v$. 因此根据命题 A.3.5 可得，或者 $x\in y$，$x=y$，或者 $y\in x$. 如果 $x\in y\in z\in U$，那么 $x\in z$，因为 z 是全的. 所以 U 在 $\in$ 关系下是线性序的. 因为 u 是全的，$x\in u\in U$，所以 U 是全的. □

定义 *一个集合 x 称为佩亚诺集，如果它是一个序数并且在关系 $\in^{-1}$ 下是良序的(所以它的每一个非空子集在关系 $\in$ 下都有一个最大元).*

结果将表明佩亚诺集是前面已证的有限序数 0，1，2，…下面先给出一些其他事实.

A.3.8 定理 *如果 x 是一个佩亚诺集且 $y\in x$，那么 y 也是一个佩亚诺集.*

证明 由命题 A.3.2 可知，y 是一个序数. 因为 x 是全的，$y\in x$，所以 y 在关系 $\in^{-1}$ 下也是良序的. □

对每一个序数 x，定义 $x+1:=x\cup\{x\}$. 由序数的定义可知，这也是一个序数.

A.3.9 命题 *如果 y 是一个序数且 x 是 y 在关系 $\in$ 下的最大元，那么 $y=x+1$.*

证明 首先，$x\subset y$，因为 y 是全的，所以 $x+1\subset y$. 反之，令 $v\in y$，$v\neq x$，那么 $v\in x$. 所以 $y\subset x+1$ 且 $y=x+1$. □

A.3.10 定理 *空集 $\varnothing$ 是一个佩亚诺集，且对任意的佩亚诺集 x，$x+1$ 也是一个佩亚诺集.*

证明 对$\varnothing$显然成立. 回想$x+1$是一个序数. 如果$A\subset x+1$，那么或者$x\in A$且x是A的最大元，或者$A\subset x$. 所以如果$A\neq\varnothing$，它就有一个最大元，因为x是一个佩亚诺集. □

A.3.11 命题 如果x是一个佩亚诺集，那么$x+1\neq\varnothing$.

证明 因为$x\in x+1$，所以结论显然成立. □

A.3.12 命题 对任意的序数x，$\bigcup(x+1)=x$.

证明 如果$y\in x$，那么因为$x\in x+1$，所以$y\in\bigcup(x+1)$. 反之，如果$t\in\bigcup(x+1)$，那么或者$t\in x$，或者$(\exists y)t\in y\in x$，但是因为x是全的，所以$t\in x$. □

A.3.13 推论 如果x和y都是佩亚诺集，且$x+1=y+1$，那么$x=y$.

A.3.14 定理(数学归纳法原理) 如果x是一个仅包含佩亚诺集的序数集，$\varnothing\in x$且对于每一个$y\in x$，都有$y+1\in x$，那么x是所有佩亚诺集的集合.

证明 首先，注意到没有包含所有序数的集合，并且由以前的经验可知，存在一个包含所有佩亚诺集的集合. 假设u是一个佩亚诺集，且$u\notin x$. 那么$u+1$作为一个序数，在关系$\in$下是良序的(定理A.3.1). 它有一个不包含在x内的最小元z. 那么$z\neq\varnothing$且由定理A.3.8和A.3.10知，z是一个佩亚诺集. 因此，z有一个最大元v. 类似地，v是一个佩亚诺集. 由z的定义，我们有$v\in x$. 根据命题A.3.9知，$z=v+1$，所以由假设知，$z\in x$，矛盾. □

现在回想定理A.2.1中的集合$\mathbb{N}$，它是满足无限公理ZF7的最小集合.

A.3.15 定理 $\mathbb{N}$是所有佩亚诺集的集合并且是一个序数.

证明 令ω是$\mathbb{N}$中所有佩亚诺集的集合. 根据定理A.2.1和A.3.14可得，ω是所有佩亚诺集的集合. 由定理A.3.10知，ω满足ZF7的条件，所以由$\mathbb{N}$的定义可得，$\omega=\mathbb{N}$. 它是一个序数集，因此在关系$\in$下是线性序的(推论A.3.6). 根据定理A.3.8可得，它是全的，所以它是一个序数. □

因此，我们得到一个无限序数ω. 然后进一步能够形成序数$\omega+1$，$(\omega+1)+1:=\omega+2$，$\omega+3$，等等. 尽管这里没有证明(事实上，它依赖于替换公理ZF8)，后面的序数序列的并是一个序数2ω. 类似地，存在并是ω^2的序数$2\omega+1$，…，3ω，…，4ω，…；并是ω^3的序数$2\omega^2$，$3\omega^2$，…；并是ω^ω的序数ω^4，ω^5，…. (注意：A^B通常表示从B映射到A的所有函数的集合，但在这里对序数不是.)这里有很多(事实上是不可数多个)可数无限序数都有相同的势ω. 存在一系列更大更复杂的序数. 这个过程没有止境，没有最大序数A，因为$A+1$仍然是一个序数. (如果存在，那么所有序数的集合将是最大的序数，这称为“Burali-Forti悖论”.)

下面将证明形式定义下的序数所具有的一些非形式性质.

A.3.16 定理 每一个良序集$(X,<)$都序同构于某一序数.

证明 对每一个$x\in X$，令$J(x):=\{y\in X:y\leqslant x\}$，$I(x):=\{y\in X:y<x\}$，$Y:=\{x\in X:J(x)$序同构于某一序数$\}$. 如果$x<y\in Y$，令$f$是$J(y)$与某一序数$B$的序同构. 那么$f\upharpoonright J(x)$是从$J(y)$的一个初始段$J(x)$到$B$的某一初始段的序同构，它是一个序数. 因此$x\in Y$，所以$Y$是$X$的一个初始段. 假设存在一个$\upharpoonright J(x)$到某一序数序同构$h$且$h\neq f\upharpoonright J(x)$. 设$t$是$X$中满足$h(t)\neq f(t)$的最小元. 但由于每个序数都是所有最小序数的集合，所以$h(t)=\{h(u):u<t\}=\{f(u):u<t\}=f(t)$，矛盾. 所以，序同构与$J(x)$是一致的且这里是唯一的. 令$f$是对所有的$y\in Y$的序同构的并. (由替换公理ZF8知，这里$f$是存在的.)那么$f$的值域(序数的一个初始段)是一个序数$B$，且$f$是一个序同构. 如果$Y=X$，我们已经证得. 否则，令$v$是$X/Y$的最小元. 但由于$f\cup\{\langle v,B\rangle\}$是一个从$J(v)$到$B+1$的一个序同构，所以$v\in Y$，矛盾. □

基数 回想称两个集合有相同的势，当且仅当存在一个从一个集合映射到另一个集合的一一对应函数. 正如每个良序集都序同构于一个且仅一个序数，基数是使得对每个集合 A，都有且仅有一个与 A 有相同势的集合. 下面是定义基数的一般方法：一个基数是一个序数，它与其任何子集都没有相同的势(更小的序数). 下面的事实表明这个定义在 1.5 节中提到的选择公理(AC)或等价地良序原理(WO)假设下都是成立的.

A.3.17 推论 *AC 或 WO 表明每一个集合 X 都与某个基数有相同的势.*

证明 由 WO，取 X 的一个良序. 由定理 A.3.16 知，取一个从 X 到某个序数 B 的序同构. 令 C 是与 B 有相同势的最小序数，g 是从 B 映射到 C 的一一对应函数. 对所有的 n'，令 $f(x) = g(h(x))$. 那么 f 是从 X 映射到基数 C 的一一对应函数. □

根据上面给出的基数为确定序数的定义，每个集合都与某一基数有相同的势这一事实蕴涵着 WO，从而蕴涵着 AC. 所以，这种处理势的方法主要依赖 AC. 由推论 A.3.6 知，任意基数集在 $\in$ 下都是良序的. 无限基数通常用带有序数下标的 $\aleph$ 表示如下：$\aleph_0 := \omega$，可数无限基数. 对每个非零序数 k，$\aleph_k$ 是大于所有 $\aleph_j (j < k)$ 的最小基数. 连续统假设表明 $2^{\aleph_0}$ 的基数为 $\aleph_1$.

由二项展开可得，存在一个从 2^ω 映上到单位区间 $[0, 1] := \{x \in \mathbb{R} : 0 \leqslant x \leqslant 1\}$ 的函数 f，其中 f 除了在一个可数集上外都是 1－1 的. 如 13.1 节中提到的，f 可被选为 1－1 的，从而使得 2^ω 和 $[0, 1]$ 具有相同的势. 它们的基数通常记为 c——“连续统”的基数. 推广的连续统假设(GCH)表明对任意无限集 X 和 Y，如果存在一个从 X 映射到 Y 和从 Y 映射到 2^X 的 1－1 函数，那么 Y 或者与 X 或者与 2^X 有相同的势. 因此，GCH 蕴涵着对每个无限基数 $\aleph_k$，它的幂集 $2^{\aleph_k}$ 与 $\aleph_{k+1}$ 有相同的势. 众所周知，GCH 蕴涵着 AC 以及 AC 既不能由通常的集合论公理(ZF1 ~ 9)证明也不能反证(参见注释).

习题

1. 证明：对任何序数 y，集合 $\{x: x \leqslant y\}$ 对于区间拓扑(在 2.2 节习题 9 中定义)是紧的.
2. 一个序数 y 称为一个序数 x 的*后继*当且仅当 $y = x + 1$，换句话说，$y = x \cup \{x\}$. 任何与其并 $\cup y$ 相等的序数称为*极限序数*(limit ordinal). 证明：一个序数要么是一个极限序数，要么是某个序数的后继，但不能两个同时成立.

A.4 从集合到数

集合论通过用集合表示各种各样的数而作为数学的基础. 证明过程的第一步，像上面那样，设置 $n + 1 := n \cup \{n\}$，用集合 $\varnothing$，$\{\varnothing\}$，$\{\varnothing, \{\varnothing\}\}$，…表示非负整数 0，1，2，…$n$ 的后继有时也写作 n'. 那么每一个非负整数 n 是含有 n 个元素的集合，记为 $\{0, 1, \cdots, n-1\}$. 现在，算术运算可通过简单的递推定义. 首先，定义加法为一个从 $\mathbb{N} \times \mathbb{N}$ 映上到 $\mathbb{N}$ 的函数 . $+(m, n)$ 写成通常的形式 $m + n$，递推定义是：$0 + n := n$，$m' + n := (m + n)'$. 这样，对每一个固定的 $n \in \mathbb{N}$，$m + n$ 是对所有的 m 通过简单的递推定义的(推论 1.3.3). 类似地，乘法也可通过简单递推定义，置 $0n := 0$，$m'n = mn + n$.

为了表示可能为负数的整数，首先定义 $\mathbb{N} \times \mathbb{N}$ 上的一个等价关系 E：$\langle m, n \rangle E \langle j, k \rangle$ 当且仅当 $m + k = j + n$. (思想是 $m - n = j - k$，且非负整数的有序对 $\langle m, n \rangle$ 是整数 $m - n$ 的一种表示.)于是对给定的等价关系，所有整数的集合 $\mathbb{Z}$ 定义为非负整数有序对的所有等价类的集合. 注意到一个非负整数将被 $\mathbb{Z}$ 中的不同集合标码，而不是先前的在 $\mathbb{N}$ 中. 然后可以利用 $\mathbb{N}$ 中的运算定义从 $\mathbb{Z} \times \mathbb{Z}$ 映上

到$\mathbb{Z}$的加法、减法和乘法. 这里不详细说明. 接下来，我们可以在有序对之间定义另一个等价关系F: $\langle u, v\rangle F\langle z, w\rangle$当且仅当$uw=vz$，其中$u$, v, z和w是$\mathbb{Z}$的所有元素，$v\neq 0$，且$w\neq 0$. 这里思想是$u/v=z/w$. 令有理数集合$\mathbb{Q}$为对F定义的有序对$<u, v>\in\mathbb{Z}\times\mathbb{Z}$ $(v\neq 0)$的所有等价类的集合. $\mathbb{Q}$中的算术运算可由$\mathbb{Z}$中的运算定义. 通常的关系$<$也可以在$\mathbb{Q}$中定义. 这里对此也不详细介绍. 现在，实数将根据戴德金分割定义. 一个集合$C\subset\mathbb{Q}$称为一个*分割*当且仅当下面的三个条件均成立：

(a) $\varnothing\neq C\neq\mathbb{Q}$.

(b) 如果$x<y\in C$，则$x\in C$.

(c) 如果$x\in C$，则$(\exists y\in C)y>x$.

所有实数的集合$\mathbb{R}$定义为所有分割的集合. 其思想为：一个实数x可用满足$q<x$的所有有理数q的集合表示. 那么实数的次序关系$\leqslant$可简单地用分割的包含关系定义. 加法定义为$C+D:=\{q+r: q\in C, r\in D\}$. 实数上的其他通常的算术运算也可以根据分割来定义，但是这里仍不给出详细情况.

实数的一个集B称为*有上界的*，当且仅当$(\exists x\in\mathbb{R})(\forall y\in B)y\leqslant x$. 这样的$x$称为$B$的一个*上界*. 如果对所有的$u<x$，$u$不是$B$的一个上界，则称$x$为$B$的*上确界*或者*最小上界*，$x:=\sup B$.

A.4.1 定理 *每一个有上界的非空实数集合B都有一个上确界.*

证明 利用分割，令$c=\cup B$. 可证明c是一个分割，且$c=\sup B$. □

类似地，实数集B称为*有下界*的当且仅当$(\exists x)$(对所有的$y\in B$, $x<y$). 一个有下界的非空实数集B有最大的下界或者下确界，记为$\inf B$.

实数可以通过无限小数展开用一种更典型的方式来定义. 但有一些困难. 例如，为了定义0.352745536… +1.647254463…，可能必须查看每一个被加数的所有数字，以得到和的第一个数字. 加法和$<$的定义及定理A.4.1的证明用分割都比用小数展开更简单. 只要实数的研究得到发展，例如在定理A.4.1中研究的实数的重要性质，那么各种用集合表示数的方式将为此起到重要作用，且也是非常有利于把实数视为直线上的点.

习题

1. 证明或反证：如果B是有下界的任意非空实数集，那么$\inf B=\cap B$. (通过分割利用实数的定义.)
2. (参考具体的ZF公理，以及简单递归、推论1.3.3)详细证明：加法是一个从$\mathbb{N}\times\mathbb{N}$映射到$\mathbb{N}$的函数.

注释

A.2节 其他等价于ZF9的公理和公理模式，称为基础公理：参见Fraenkel et al(1973, p. 86—91). 对集合论研究更多的另一个公理体系是Von Neumann-Gödel-Bernays(GB)体系. 相比于ZF，在GB中，不仅有集合，而且有称为类的对象. 所有的集合都是类. 一些类不是集合，称其为*真类*(proper class). 类A是一个集合当且仅当对某个类B, $A\in B$，所以在GB中定义"集合"："类"和"$\in$"是GB的基本的未定义概念. 公理ZF1、ZF2、ZF3、ZF5、ZF6、ZF7和ZF9也是GB中的公理(或定理)，且ZF1被假定为类和集合. 代替ZF4，GB有一个有限多个(8或9)公理的集合，其全体等价于公理模式. 公理模式对只有一个自由变量x的任意合式公式P，存在一个满足$x\in A$的类A当且仅当P. 代替公理模式ZF8，GB有一个公理：

$$\forall A(\forall \text{ 集合 } t)(((\forall x\in t)\forall y\forall z((\langle x,y\rangle\in A \quad \text{且} \quad \langle x,z\rangle\in A\Rightarrow y=z))\Rightarrow$$

$$(\exists \text{ 集合 } v)(\forall y)(y \in v \text{ 当且仅当}(\exists x \in t)\langle x,y\rangle \in A))).$$

因此，GB 有一个公理和无公理模式的有限集合．它允许大的对象被命名，比如‘全域 V’，所有集合的类．（一般来说，真类是那些太大而不能称之为集合的类．）大类产生于关于范畴理论（同调代数）的研究．暂时，GB 好像比 ZF 提供了一个更方便的范畴理论基础．但是范畴理论伴随着全域 V 有时候也在 GB 之上成立，考虑对象（如 2^V）（参见 MacLane，1971.）

集合有序对的表示归功于 Wiener（1914）和 Kuratowski（1921）．Wiener（1894—1964），以其在其他领域的工作而闻名，在 19 岁的时候用“类型论”语言写了一篇文章．Bertrand Russell 和 Alfred North Whitehead 发展了这一理论，但没有再应用过．Kuratowski 在波兰杂志《Fundamenta Mathematicae》第二卷发表了他的论文，不久成为该杂志的主编，直到他 84 岁去世（参见 Fundamenta，1980）.

参考文献中还包括了很多关于集合论的书．Bishop（1967）对分析提出了一个十分不同的基础，没有任何精确的公理体系．关于 ZF 和 GB 的关系参见 Cohen（1966，p. 73—78）．Quine（1951，p. 33—37）在他文章的第二部分关于“拟-引用”提出了“角标”.

A. 3 节 序数的发展是基于 Kelly（1955）的．他把序数的定义归功于 Raphael M. Robinson.

Dedekind（1888）和 Peano（1889，1891）对非负整数根据 0 和后续运算给出了一个公理集合．根据定理 A. 3. 10、命题 A. 3. 11、推论 A. 3. 13、定理 A. 3. 14 和定理 A. 3. 15，这些公理对于 $0:=\varnothing$ 和 $x':=x\cup\{x\}$ 成立（参见 Kleene，1952，p. 20）．尽管我没有像在文献中那样看到“佩亚诺集”的术语，但看起来这个术语是有用的.

Gödel（1940）证明了如果 ZF1 ~ ZF9 是一致的，那么 GCH（CH 也是）和 AC 与 ZF1 ~ ZF9 是一致的．换句话说，如果 GCH 与 ZF1 ~ ZF9 矛盾，那么仅从 ZF1 ~ ZF9 就可得出矛盾．集合论强烈认为 ZF1 ~ ZF9 是一致的．Paul Cohen（1963—1964，1966）证明了 CH 独立于 ZF1 ~ ZF9 和 AC．他建立了一个模型（model），此模型是对 AC 和 ZF1 ~ ZF9 成立，但在此情况下对 CH 不成立的集族：特殊地，$2^{\aleph_0}$ 的基数是 $\aleph_2$．他还通过对 ZF1 ~ ZF9 建立模型证明了 AC 与 ZF1 ~ ZF9 独立，在 ZF1 ~ ZF9 中 AC 不成立，甚至对有两个元素的可数个集合 A_i 的一个笛卡儿积，或者在不是良序的 $\mathbb{R}$ 中（Cohen，1966，p. 138）．相反地，在对某个 x_i 和 y_i，每一个 $A_i=\{x_i, y_i\}=\{y_i, x_i\}$ 的模型中，$A:=\bigcup_{i=1}^{\infty} A_i$ 是不可数的：不存在从 $\mathbb{N}$ 映上到 A 的函数.

考虑 $\mathbb{N}$ 上的所有良序集的集合 S，令 S/I 是在序同构下 S 中元素的所有等价类的集合，那么 S/I 是基数为 $\aleph_1$ 的集合．不存在一个从 S/I 映上到 $[0, 1]$ 上的已知函数，且的确不可能存在，因为 CH 是独立的．在这个意义下，一些集合论学家认为，CH 在“现实世界”中是错误的．不过，它仍然是一个有用的、一致的和相当受欢迎的假设．另一方面，由于 S/I 是良序集，CH 蕴涵着 $[0, 1]$ 可能是良序的，所以对 $[0, 1]$ 的子集蕴涵着选择公理．这是 Sierpiński（1947）证明的 GCH 蕴涵着 AC 的一个例子.

A. 4 节 Landau（1930）用一个满足佩亚诺公理的后继运算给出了一个从正整数到实数的详细讨论．J. Van Hijenoort（1976）就数学基础提供了其他的文献.

在“非标准分析”中，存在一个无限小但并不为 0 的实数．例如，参见 Nelson（1977）和 Robinson（1979）．通常分析中的一些事实能够用非标准分析技术更明显地证明.

参考文献

Bishop, Errett (1967). *Foundations of Constructive Analysis*. McGraw-Hill, New York.

Church, Alonzo (1956). *Introduction to Mathematical Logic, I*. Princeton University Press.

Cohen, Paul J. (1963–1964). The independence of the continuum hypothesis. *Proc. Nat. Acad. Sci. USA 50*: 1143–1148, *51*: 105–110.

——— (1966). *Set Theory and the Continuum Hypothesis*. Benjamin, New York.

Dedekind, Richard (1888). *Was sind und was sollen die Zahlen*? Braunschweig.

Fraenkel, Abraham A., Yehoshua Bar-Hillel, and Azriel Levy (1973). *Foundations of Set Theory*. 2d ed. North-Holland, Amsterdam.

Fundamenta Mathematicae 109 (1980). Obituary of K. Kuratowski. pp. i–ii.

Gödel, Kurt (1940). *The Consistency of the Continuum Hypothesis. Ann. Math. Studies*, no. 3. Princeton University Press.

van Heijenoort, Jean (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press.

Kelley, John L. (1955). *General Topology*. Van Nostrand, Princeton.

Kleene, Stephen C. (1952). *Introduction to Metamathematics*. Van Nostrand, Princeton.

——— (1967). *Mathematical Logic*. Wiley, New York.

Krivine, Jean-Louis (1971). *Introduction to Axiomatic Set Theory*. Reidel, Dordrecht.

Kuratowski, Kazimierz (1921). Sur la notion d'ordre dans la théorie des ensembles. *Fund. Math.* 2: 161–171.

Landau, Edmund (1930). *Foundations of Analysis*. Transl. Chelsea, New York, 1951.

MacLane, Saunders (1971). *Categories for the Working Mathematician*. Springer, New York.

Nelson, Edward (1977). Internal set theory: A new approach to nonstandard analysis. *Bull. Amer. Math. Soc. 83*: 1165–1198.

Peano, Giuseppe (1889). *Arithmetices principia, nova methodo exposita*. Turin. Also in *Opere scelte*, II, pp. 20–55 (Edizioni Cremonese, Rome, 1958); transl. in van Heijenoort (1967), pp. 85–97.

——— (1891). Sul concetto di numero. *Rivista di Matematica 1*: 87–102, 256–267.

Quine, Willard Van Orman (1951). *Mathematical Logic*. 2d ed. Harvard University Press.

Robinson, Abraham (1979). *Selected Papers of Abraham Robinson*. Vol. II, *Nonstandard Analysis and Philosophy*. Edited and with introductions by W. A. J. Luxemburg and S. Körner. Yale University Press.

Sierpiński, Wacław (1934, 1957). *Hypothèse du continu*. Monografje Matematyczne, Warsaw. 2d ed., "Containing as Appendix a number of research papers, 1934–1956." Chelsea, New York.

——— (1947). L'hypothèse généralisée du continu et l'axiome du choix. *Fundamenta Mathematicae 34*: 1–5. Repr. in Sierpiński (1957).

Struik, Dirk J. (1948). *A Concise History of Mathematics*. 2d ed. Dover, New York.

Suppes, Patrick (1960). *Axiomatic Set Theory*. Van Nostrand, Princeton.

Wiener, Norbert (1914). A simplification of the logic of relations. *Proc. Camb. Philos. Soc. 17*: 387–390.

附录 B　复数、向量空间和泰勒余项定理

本章从实数开始定义复数.（正如戴德金分割，实数在 1.1 节和附录 A.4 中定义.）复数域$\mathbb{C}$定义为具有通常加法

$$(x,y)+(u,v):=(x+u,y+v)$$

和复数乘法

$$(x,y)(u,v):=(xu-yv,xv+yu)$$

的平面$\mathbb{R}^2$. 把实直线$\mathbb{R}$视为$\mathbb{C}$的子集，令 $x:=(x,0)$，认为实数 x 与$\mathbb{C}$中$(x,0)$相对应. 从而$\mathbb{R}\times\mathbb{R}$上的加法和乘法与$\mathbb{C}\times\mathbb{C}$限制到$\mathbb{R}\times\mathbb{R}$上的加法和乘法是一致的.

复数 i 定义为$(0,1)$. 因此 $i^2=(-1,0)=-1$，并且$(x,y)=x+iy$；复数一般都写成后一种形式. 从而它们的加法和乘法就可由实数的加法和乘法及 $i^2=-1$ 来确定.

对任意复数 $z=x+iy$，复共轭定义为 $\bar{z}:=x-iy$. 绝对值定义为 $|z|:=(x^2+y^2)^{1/2}$. 对任意的复数 $z\neq0$，令

$$1/z:=\bar{z}/|z|^2=(x/(x^2+y^2),-y/(x^2+y^2)).$$

那么 $z(1/z)=1$. 复数 w 与 $z(z\neq0)$的除法定义为

$$w/z:=w(1/z).$$

如果 $z=x+iy$，其中 x 和 y 是实数，那么 x 称为实部(real part)，记为 $x=\mathrm{Re}(z)$，y 称为 z 的虚部(imaginary part)，记为 $y=\mathrm{Im}(z)$. 对实的 y，数 iy 称为纯虚部(purely imaginary).

现在向量空间将定义在域 K 上，本书中 $K=\mathbb{R}$ 或者 $K=\mathbb{C}$. K 上的一个向量空间 V 是通过从 $V\times V$ 映射到 V 的"加法"$(v,w)\rightarrow v+w$ 和从 $K\times V$ 映射到 V 的"数量乘法"$(c,v)\mapsto cv$ 定义的，且加法和数量乘法具有如下性质：$(V,+)$是一个阿贝尔群，意味着对 V 中所有的 u，v 和 w，$u+v=v+u$，$u+(v+w)=(u+v)+w$，存在一个 $0\in V$，使得对所有的 u，都有 $u+0=u$，和对任意的 $v\in V$，存在某个 $-v\in V$，使得 $v+(-v)=0$. 同样，数量乘法也要求满足对任意的 $v\in V$ 和 a，$b\in K$，$(ab)v=a(bv)$，$1\cdot v=v$，$(-1)\cdot v=-v$，以及$(a+b)v=av+bv$. 依然有 $0\cdot v=0$. 这样定义的向量空间有时也称为线性空间(linear space).

开区间(a,b)上的一个实值函数 f 称为 C^n 的，如果对 $j=1,2\cdots,n$，导数 $f^{(j)}$ 存在，其中 $f^{(0)}\equiv f$，$f^{(1)}\equiv f'$，且 $f^{(n)}$ 在整个区间(a,b)上是连续的(那么低阶导数自然也是连续的). 假设 $0\in(a,b)$，那么可以定义 f 在 0 点附近的一个部分泰勒级数展开式，且余项为

B.1
$$R_n(f,x):=f(x)-\sum_{j=0}^{n-1}f^{(j)}(0)x^j/j!,$$

其中 $a<x<b$. 从而我们有下面的结论.

B.2 具有积分型余项的泰勒定理　如果 f 在(a,b)上是 C^n 的，且 $a<0<b$，那么

B.3
$$R_n(f,x)=\int_0^x f^{(n)}(t)(x-t)^{(n-1)}\mathrm{d}t/(n-1)!.$$

证明　当 $n=1$ 时，由微积分基本定理可得(B.3)成立. 一般情况可以通过对 n 归纳证得. 如果 $n\geq2$，那么归纳假设给出

$$R_{n-1}(f,x) = -\int_0^x f^{(n-1)}(t)\mathrm{d}_t(x-t)^{n-1}/(n-1)!,$$

其中 d_t 表明 t 是积分变量，x 是固定的. 由分部积分可得，

$$\frac{f^{(n-1)}(0)x^{n-1}}{(n-1)!} + \int_0^x f^{(n)}(t)(x-t)^{n-1}\mathrm{d}t/(n-1)!,$$

由此推出(B.3). □

B.4 推论 在相同的条件下，

$$|R_n(f,x)| \leqslant \sup_{0\leqslant t\leqslant 1}|f^{(n)}(tx)||x|^n/n!.$$

现在考虑可能的复值函数. 具有复值项 z_n 的级数 $\sum_{n=1}^{\infty} z_n$ 称为绝对收敛的(absolutely convergent)，如果 $\sum_n |z_n| < \infty$. 容易看出一个绝对收敛的级数在$\mathbb{C}$上关于通常度量 $d(z, w) := |z-w|$ 是收敛的. 如果函数 f 在区间$(a-r, a+r)$上有收敛的泰勒级数

$$f(x) = \sum_{n=0}^{\infty} f^{(n)}(a)(x-a)^n/n!,$$

那么对圆 $|z-a| < r$ 上的所有复数 z，$f(z)$由如下相同的级数定义：

$$f(z) := \sum_{n=0}^{\infty} f^{(n)}(a)(z-a)^a/n!,$$

它对每一个这样的 z 都是绝对收敛的. 特别地，复数指数函数定义为 $e^z := \sum_{n=0}^{\infty} z^n/n!$，它对所有的复数 z 都是绝对收敛的. 对任意的实数 t，从已知的正弦和余弦函数的泰勒级数可得 $e^{it} = \cos t + i\sin t$，所以 $|e^{it}| = 1$. 由二项式定理和绝对收敛知，对任意的复数 w，z，乘法法则 $e^{w+z} = e^w + e^z$ 仍然成立. 容易看出对任意的实数 x，y，$|e^{x+iy}| = e^x$. 对 w，$z\in\mathbb{C}$，$e^w = e^z$ 当且仅当对某个 $m\in\mathbb{Z}$，$z = w + 2m\pi i$.

任意的 $z\in\mathbb{C}$，除了 $z=0$，都可以唯一地写成 $z = re^{i\theta}$，其中 $r\geqslant 0$，$\theta\in\mathbb{R}$ 且 $-\pi\leqslant\theta\leqslant\pi$. 从而对数的主支(principal branch of the logarithm)定义为 $\mathrm{plog}(z) := \log r + i\theta$，其中 $r>0$，$r=0$ 时没有定义，这里对 $r>0$，$\log r$ 是通常的实值对数. 函数 plog 对 $|\theta| < \pi$ 是解析的，但是当 $|\theta| = \pi$ 时是不连续的. 因此有 $\mathrm{plog}(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \cdots$，其中 $|z| < 1$. 对 w，$z\in\mathbb{C}$，我们有 $w = e^z$ 当且仅当对某个 $m\in\mathbb{Z}$，$z = \mathrm{plog}(w) + 2m\pi i$.

如果复数导数存在，则定义为 $f'(z) := \lim_{w\to 0}(f(z+w) - f(z))/w$. 从复平面$\mathbb{C}$的一个开子集 U 映射到$\mathbb{C}$的一个函数 f 称为全纯的(holomorphic)或解析的，当且仅当对所有的 $z\in U$，导数 $f'(z)\in\mathbb{C}$ 存在. 如果 f 在 U 上是全纯的，那么它的各阶导数在 U 上存在. 对每个 $w\in U$，存在一个最大的 $r(0<r\leqslant+\infty)$. 使得 $|z-w| < r$，这表明 $z\in U$，并且对任意这样的 w 和 z，$f(z)$在 w 附近的泰勒级数收敛：

$$f(z) = \sum_{n=0}^{\infty}\frac{f^{(n)}(w)}{n!}(z-w)^n$$

(Ahlfors，1979，p.38，119，179). 现在在(B.1)、(B.2)、(B.3)和(B.4)中，实值变量 x 都可以用复变量 z 代替，如果在(B.2)中，f 在开圆 $D := \{z: |z| < r\}(z\in D)$上是全纯的，(B.3)和(B.4)

中的积分变量 t 被复变量 w 取代，沿线段从 0 到 z 积分. Marsden（1973，p. 83）给出了复数情况下的微积分基本定理.

推论 B. 4 对小的 $|w|$ 或 $|z|$ 最有用. 对大的 $|z|$ $(z\in D)$，推论可能不太有用，正如下面 4 个例子所说明的.

例(a)：令 $f(z):=1/(1-z)=1+z+z^2+\cdots$，$|z|<1$(几何级数). 那么 $f^{(n)}(z)=n!/(1-z)^{n+1}$. 因此对 $0<x<1$，推论 B. 4 仅给出 $|R_n(f,\ x)|\leqslant x^n/(1-x)^{n+1}$，且仅对 $x<1/2$ 有当 $n\to\infty$ 时它趋于 0.

因此，在最大开圆内无穷泰勒级数的收敛性不能从推论 B. 4 得出，而且也不能很容易地从(B. 3)得出. 在此最大开圆内函数是全纯的，这可由柯西积分公式证得(Ahlfors，1979，p. 38，119，179).

例(b)：令 $f(z)=\operatorname{plog}(1+z)$，那么 $f^{(n)}(z)=(-1)^{n-1}(n-1)!/(1+z)^n$，$n=1,\ 2,\ \cdots$，推论 B. 4 给出 $|R_n(f,\ z)|\leqslant|z|^n/(n(1-|z|)^n)$，仅当 $|z|\leqslant 1/2$，$n\to\infty$ 时，它趋于 0.

例(c)：回想对 $x\in\mathbb{R}$，$k=1,\ 2,\ \cdots$，$\binom{x}{k}:=x(x-1)\cdots(x-k+1)/k!$，而 $\binom{x}{0}:=1$. 对满足 $|t|<1$ 的实数，由 $g(t):=(1-t)^{1/2}$ 定义的函数 g 有泰勒级数 $f(t):=\sum_{n=0}^{\infty}\binom{1/2}{n}(-t)^n$. 比较分子和分母中的因子，我们看到对所有的 $n=0,\ 1,\ \cdots$，$\left|\binom{1/2}{n}\right|<1$，从而对 $|t|\leqslant r<1$，级数 $f(t)$ 一致且绝对收敛，并且定义了 $f(t)$. 由定理 B. 2 得到，至少对 $|t|<1/2$，$f(t)=(1-t)^{1/2}$. 同样，二重级数

$$f(t)^2=\sum_{j,m=0}^{\infty}\binom{1/2}{j}\binom{1/2}{m}(-t)^{j+m}$$

对 $|t|\leqslant r$ 一致且绝对收敛，所以能够被重排为 $\sum_{k=0}^{\infty}C_kt^k$，其中

$$C_k:=(-1)^k\sum_{i=0}^{k}\binom{1/2}{i}\binom{1/2}{k-i}.$$

当 $|t|<1/2$ 时，$f(t)=g(t)=(1-t)^{1/2}$，因此可以得出 $C_0=1$，$C_1=-1$ 和 $C_k=0(k\geqslant 2)$. 所以对 $|t|<1$，$f(t)=g(t)=(1-t)^{1/2}$(由连续性知，它不可能是 $-(1-t)^{1/2}$)，且对 $|t|\leqslant r$，级数 $f(t)$ 一致收敛到 $g(t)$.

对 $(1+x)^a$，其中 a 是任意实数，Courant(1937)给出了二项式级数收敛性的另一个基本证明，附录的第Ⅵ章，第 3 节.

例(d)：对 $x\neq 0$，令 $f(t):=e^{-1/x^2}$，$f(0):=0$. 那么很容易检验出 f 是 C^∞ 的，即它具有连续的各阶导数. 它的泰勒级数在 0 附近的系数都为 0，所以级数几乎处处收敛，但是除了 $x=0$ 外，不收敛到 $f(x)$.

下面将对一元实函数说明关于带有余项的泰勒级数的另一个事实. 回想 $f=o(g)(t\to 0)$ 意味着 $f/g\to 0$，同时 $f=O(g)$ 意味着 f/g 是有界的.

B. 5 定理 如果 $n\geqslant 1$，f 是一个含 0 的区间上的实值函数，使得 n 阶导数 $f^{(n)}(0)$ 是有定义且有限的，那么 $R_{n+1}(f,\ t)=o(t^n)(t\to 0)$.

证明 由假设，在0的某一邻域内对t的导数$f^{(j)}(t)(j=1, 2, \cdots, n-1)$存在. 从函数$f$减去一个多项式，我们可以假设$f^{(j)}(0)=0$，其中$j=0, 1, \cdots, n$. 如果对某个函数$h$和非负不减函数$k$，当$t\to 0$且$h(0)=0$时，$h'(t)=O(k(t))$，那么由中值定理知，$h(t)=O(tk(t))(t\to 0)$. 类似地，如果$h'(t)=o(k(t))$，则有$h(t)=o(tk(t))$. 由导数的定义知，$f^{(n-1)}(t)=o(t)$，从而通过迭代就可得$f(t)=o(t^n)$. □

参考文献

Ahlfors, Lars Valerian (1979). *Complex Analysis*. 3d ed. McGraw-Hill, New York.

Courant, R. (1937). *Differential and Integral Calculus*, Vol. I, 2d ed., transl. by E. J. McShane. Interscience, New York.

Marsden, Jerrold E. (1973). *Basic Complex Analysis*. W. H. Freeman, San Francisco.



附录C 测度问题

正如3.4节所陈述的，假设选择公理，则对勒贝格测度存在一个不可测集. 还有一个问题：勒贝格测度作为一个可数可加的测度是否可以扩张到$\mathbb{R}$的所有子集上. 更一般地，“测度问题”问：在一个不可数集X的所有子集上是否存在一个测度μ，使得$\mu(X)=1$且对每一个$x\in X$，$\mu\{x\}=0$. 附录C将会给出部分答案，它将会在附录E中用到.

C.1 定理(Banach 和 Kuratowski) *在连续统假设下，不存在定义在$I:=[0,1]$的所有子集上的测度μ，使得$\mu(I)=1$，对每一个$x\in I$，$\mu\{x\}=0$.*

证明 证明将基于下面的引理.

C.2 引理 *在连续统假设下，存在具有下列性质的I的子集A_{ij}，其中i，$j=1$，$2\cdots$：*

(a)对每个i，集合A_{ij}对不同的j是不相交的，并且其并是I.

(b)对任意的正整数序列$k(i)$，

$$\bigcap_i \bigcup_{1\leq j\leq k(i)} A_{ij} \text{ 至多是可数的.}$$

证明 对任意的两个正整数序列$\{n_i\}$和$\{k_i\}$，$\{n_i\}\leqslant\{k_i\}$意味着对所有的i，$n_i\leqslant k_i$. 我们先证明下面的引理.

C.3 引理 *在连续统假设下，存在一个整数序列的集合$\mathcal{F}$，$\mathcal{F}$的势为c(I的势)，使得对每一个序列$\{m_j\}$(在$\mathcal{F}$中或不在$\mathcal{F}$中)，$\mathcal{F}$中满足$\{n_j\}\leqslant\{m_j\}$的所有序列$\{n_j\}$的集合至多是可数的.*

证明 令$\mathcal{S}$是所有正整数序列的集合. 那么$\mathcal{S}$的势为c(13.1节). 由连续统假设知，在$\mathcal{S}$中存在一个良序$\leqq$，使得对每个$y\in\mathcal{S}$，$\{x:x\leqq y\}$是可数的. 对每个$\alpha\in\mathcal{S}$，令f_α是从$\mathbb{N}$映上到$\{x:x\leqq\alpha\}$的函数. 由$g_\alpha(n):=f_\alpha(n)(n)+1(n=0,1,\cdots)$定义一个正整数序列$g_\alpha:=\{g_\alpha(n)\}_{n\geqslant 0}$. (那么$g_\alpha$称为“对角线”序列.)对任意可数多个$x\leqq\alpha$来说，$g_\alpha\leqslant x$是不成立的. 令$\mathcal{F}$是所有序列$g_\alpha(\alpha\in\mathcal{S})$的集合. 那么如果$g_\alpha\leqslant x$，一定不是$x\leqq\alpha$的情况，所以$\alpha\leqq x$，并且这种$\alpha$的集合是可数的. 现在$\mathcal{F}$是不可数的，因为对于某个$y\in\mathcal{S}$，每个$g_\alpha$是序列$y$，如果这种$y$的集合是可数的，那么对$\leqq$，它们将有一个上确界$\beta\in\mathcal{S}$，但是序列$g_\beta$与所有$y\leqq\beta$的序列都不相同，矛盾. 所以根据连续统假设，$\mathcal{F}$的势为$c$，引理C.3得证. □

现在来证明引理C.2，令h是从$[0,1]$映上到$\mathcal{F}$的$1-1$函数. 因此，每个$h(x)$都是一个序列$\{h(x)_n\}_{n\geqslant 0}$. 由$x\in A_{ij}$定义集合$A_{ij}$当且仅当$j=h(x)_i$. 那么对一个固定的$i$，集合$A_{ij}$对不同的$j$显然是不相交的. 它们对所有$j$的并就是整个区间$[0,1]$.

令$\{k_i\}:=\{k(i)\}$为任意正整数序列，$x\in\bigcap_i\bigcup_{j\leqslant k(i)}A_{ij}$. 那么由$A_{ij}$的定义，对所有的$i$，我们有$h(x)_i\leqslant k(i)$，所以$h(x)\leqslant\{k_i\}$. 由引理C.3，在$\mathcal{F}$中仅有可数多个序列，它们小于等于$\{k_i\}$，且由于$h$是$1-1$的，仅存在可数多个$x\in[0,1]$，引理C.2得证. □

现在为了证明定理C.1，对每个$i\geqslant 1$，取$k(i)$，使得$\mu(B_i)<1/2^{i+1}$，其中$B_i:=\bigcup_{j>k(i)}A_{ij}$. 根据引理C.2可知，所有$B_i$的补集的交是可数的，所以它的$\mu$测度为0. 因此$1=\mu(\bigcup_i B_i)<\sum_{i\geqslant 1}1/2^{i+1}=1/2$，矛盾. □

注释

这个附录是基于 Banach 和 Kuratowsky(1929) 的论文给出的. 关于测度问题的更多信息("可测基数"、"实值可测基数"等.), 参见 Jech (1978).

参考文献

Banach, Stefan, and Casimir [Kazimierz] Kuratowski (1929). Sur une généralisation du problème de la mesure. *Fund. Math. 14*: 127–131.

Jech, Thomas (1978). *Set Theory*. Academic Press, New York.



附录 D　非负项的重排和

非负项的和可以重新排列是分析学中的一个基本事实(例如，Stromberg，1981，p. 61)．这里将给出完备性证明．

D. 1 引理　假设对所有的 $m, n\in\mathbb{N}$，$a_{mn}\geqslant 0$，令 $k\mapsto(m(k), n(k))$ 是从 $\mathbb{N}$ 映射到 $\mathbb{N}\times\mathbb{N}$ 的任意一个一一对应函数，那么

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}a_{mn}=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}a_{mn}=\sum_{k=0}^{\infty}a_{m(k),n(k)}=S$$
$$:=\sup\Big\{\sum_{\langle m,n\rangle\in F}a_{mn}:F\subset\mathbb{N}\times\mathbb{N},F\text{ 有限}\Big\}$$

(无论 S 是有限的还是 $+\infty$)．

证明　对有限部分和，我们有

$$S_{MN}:=\sum_{m=0}^{M}\sum_{n=0}^{N}a_{mn}=\sum_{n=0}^{N}\sum_{m=0}^{M}a_{mn}\leqslant S,$$
$$T_K:=\sum_{k=0}^{K}a_{m(k),n(k)}\leqslant S,$$

其中 M，N 和 K 为任意正整数．因此可以用 $+\infty$ 来替换每一个和中的上限 M，N 和 K(首先内层和，然后外层和)，且它们保持小于或等于 S．另一方面，对任意有限的 $F\subset\mathbb{N}\times\mathbb{N}$，存在某个有限的 K 和 M，满足

$$S_F:=\sum_{\langle m,n\rangle\in F}a_{mn}\leqslant T_K\leqslant S_{MM}.$$

因此 $S_F\leqslant T_\infty$，$S_F\leqslant\lim\limits_{M\to\infty}\lim\limits_{N\to\infty}S_{MN}:=U$ 且 $S_F\leqslant\lim\limits_{N\to\infty}\lim\limits_{M\to\infty}S_{MN}:=V$．从而 $S\leqslant T_\infty$，$S\leqslant U$ 且 $S\leqslant V$，所以 $S=T_\infty=U=V$．　□

参考文献

Stromberg, Karl R. (1981). *An Introduction to Classical Real Analysis*. Wadsworth, Belmont, Calif.

附录 E　非度量紧空间的病态性

回想一个拓扑空间称为*局部紧的*，当且仅当它中的每一个点都包含在某一具有紧闭包的开集中. 在某些情况下，局部紧空间被认为是研究一般测度和积分的最自然的空间. 正如 7.3 节中所讨论的，正则性扩张在紧或局部紧空间上是作为测度的一个最主要优点出现的. 局部紧性也很适用于群结构. *拓扑群*是具有豪斯多夫拓扑$\mathcal{T}$的一个群 G，对此拓扑，群运算$\langle g, h\rangle \mapsto gh$ 和$g \mapsto g^{-1}$是连续的. 一个局部紧的豪斯多夫拓扑群具有左和右的*哈尔测度*(Haar measure)，它们分别是在所有左或右平移 $g\mapsto hg$ 或 $g\mapsto gh$ 下的拉东测度(在紧集上有限)不变量. 关于局部紧空间上的测度论，尤其是群，占了 Helmos(1950)经典教材 12 章内容的最后三章. Bourbaki(1952—1969)甚至把更多的重点放在了局部紧空间上. 本附录的主要目的是指出为什么局部紧空间在本书中比在 Helmos(1950)和 Bourbaki(1952—1969)所占篇幅少.

测度论的另一类空间是完备可分度量空间，或能够度量成可分的且完备的拓扑空间(波兰空间). 在这种空间中，例如，可分的无穷维巴拿赫空间，它们当中没有一个是局部紧的. 但是在这种空间中，由于 Ulam 定理(7.1.4)，局部紧的很多优点仍然保持，至少对有限测度(比如概率测度).

如果测度论主要在局部紧空间中研究，正如在 Bourbaki 的教材中，那么一个非常一般的(完全正则的豪斯多夫)拓扑空间可以被认为是一个紧豪斯多夫空间的子集(定理 2.8.3). 然而，其他的结构(比如代数运算(加法，标量乘法))不能连续地扩张到紧化，并且吉洪诺夫-切赫的一般紧化通常也会丢失其他的性质，比如可度量化(参见 2.8 节中的习题 5 和习题 7).

或许更严重的是，对困难认识不足制约着非度量紧豪斯多夫空间的测度论的发展. 例如，一个从单位区间[0, 1]映射到一个紧豪斯多夫空间上的连续函数序列f_n可能处处收敛于一个函数f，而f是不可测的(对值域空间上的博雷尔 σ-代数)，正如命题 4.2.3 所示. 本附录中的主要事实和随机过程的结构有关，正如在科尔莫戈罗夫定理(12.1.2)中所表明的. 令 T 是任意一个集合，$\mathbb{R}^T$ 是 T 上所有实值函数的集合，如果由科尔莫戈罗夫定理给定$\mathbb{R}^T$上的一个概率测度，那么它只在最小σ-代数上有定义，对最小 σ-代数，坐标值 e_t，$e_t(f) := f(t)$ 对每个 $t\in T$ 都是可测的. 那么一个可测集只依赖于可数个坐标. 因此，如果 T 是不可数的，那么甚至单元素集$\{f\}$也不是可测的，并且f在 T 上的上确界也不是f的可测函数. 所以有必要把过程的“法则”扩张到一个较大的σ-代数上. 这是可以做到的，例如，对维纳过程(布朗运动)，通过样本连续性，根据定理 12.1.5，在连续函数空间上定义法则. 更一般地，对满足 $T\subset\mathbb{R}$ 的合适过程，可以在具有左极限的右连续函数的集合上定义它们的法则，正如下面的定理 E.6 所述.

7.3 节中的正则性扩张可以应用到实值过程.

E.1 构造　对任意的实值随机过程 $x(t, w)$，$t\in T$，$w\in\Omega$，我们可以认为 $x(\cdot, \cdot)$在紧可度量空间$\overline{\mathbb{R}} := [-\infty, +\infty]$上取值. 在它的博雷尔 σ-代数下，$\overline{\mathbb{R}}$ 是一个标准可测空间(12.1 节). 科尔莫戈罗夫定理(12.2.1)给出了一个在紧的积空间$\overline{\mathbb{R}}^T$ 上 x 的法则 P_x，所以对任意有限集 $F=\{t_1, \cdots, t_k\}\subset T$ 和博雷尔集 $A_i\subset\mathbb{R}$ ($i=1, \cdots, k$)或$\overline{\mathbb{R}}$，有

$$P\{x(t_i, \cdot)\in A_i, i=1,\cdots,k\} = P_x\{f\in\overline{\mathbb{R}}^T : f(t_i)\in A_i, i=1,\cdots,k\}.$$

P_x 定义在最小 σ-代数$\mathcal{C}$上，对此最小 σ-代数$\mathcal{C}$，在$\overline{\mathbb{R}}^T$ 上所有的坐标投影 π_t 都是可测的，其中对每个 $f\in\overline{\mathbb{R}}^T$，$\pi_t(f):=f(t)$．正如在定理 7.1.1 后面的例子中所看到的，在紧豪斯多夫空间$\overline{\mathbb{R}}^T$ 中，$\mathcal{C}$等于贝尔 σ-代数．因此，根据正则性扩张(7.3 节)，P_x 在$\overline{\mathbb{R}}^T$ 上扩张到一个正则博雷尔测度 $P_{(x)}$．

然而，注意到这个构造使得 T 上任意结构都没有用，比如一个序、拓扑或者 σ-代数．但是，假设$\mathcal{S}$是 T 的子集的一个 σ-代数，$(\Omega,\ \mathcal{A},\ P)$是一个概率空间，$x$ 是一个可测的随机过程，即一个从 $\Omega\times T$ 映射到$\mathbb{R}$ 的联合可测函数．

对某个积 σ-代数的完备化，从$\overline{\mathbb{R}}^T\times T$ 映射到$\mathbb{R}$ 的函数$(f,\ t)\mapsto f(t)$是可测的吗？命题 E.2 将给出一个否定的回答，甚至对一个类正态过程(正如在 12.1 节所定义的)．这表明紧化和正则性扩张不能产生想要的结果．但另一方面，如果 $T=[0,\ 1]$，过程 x 处处具有左、右极限，那么在博雷尔集合上具有这种极限的函数是可测的且正则性也可以得到拓展(参见定理 E.6)．

令 H 是一个可分的无穷维希尔伯特空间，具有标准正交基$\{e_n\}_{n\geq1}$．那么对每个 $y\in H$，对某个满足 $\sum_n y_n^2<\infty$ 的实数 y_n，$y=\sum_n y_ne_n$．令 G_n 是定义在概率空间$(\Omega,\ \mathcal{A},\ P)$上的独立同分布随机变量，服从标准正态分布 $N(0,\ 1)$．具体地，令 Ω 是具有坐标 w_n 的实直线的一个可数笛卡儿积，其中对每个 n，P 是$\mathbb{R}$ 上的标准正态法则 $N(0,\ 1)$的积．正如在 12.1 节中，这个类正态过程可以由 $L(y)(w):=\sum_n y_nw_n$ 定义．对于每个 $y\in H$，由三级数定理(9.7.3)知，这个级数几乎必然收敛．如果级数不收敛，则令 $L(y)(w):=0$．定义 L 的级数的有限部分和显然在 $H\times\Omega$ 上是联合可测的，且在 H 上具有博雷尔 σ-代数．可测函数序列(或级数)收敛的集合是可测的，正如定理 4.2.5 所证明的那样．所以所定义的 L 是联合可测的．

由构造 E.1 知，H 上的随机过程 L 定义了贝尔 σ-代数$\overline{\mathbb{R}}^H$ 上的一个概率测度 P_L 和其正则性扩张 $P_{(L)}$，$P_{(L)}$在$\overline{\mathbb{R}}^H$ 中的博雷尔集 σ-代数上作为测度是有定义和正则的．从而 $P_{(L)}$有一个完备化 $\bar{P}_L$(命题 3.3.3)，所以只要 $A\subset B$ 且 $P_{(L)}(B)=0$，那么 $\bar{P}_L(A)=0$．

对$\overline{\mathbb{R}}^H\times H$ 上的一个积 σ-代数，函数$(f,\ t)\mapsto f(t)$不是联合可测的，即使$\overline{\mathbb{R}}^H$ 上的 σ-代数是所有它自身子集的集族，只要至少一个集合 $A\subset H$ 不在 H 上所考虑的 σ-代数中(令 $f=1_A$)．所以为了得到联合可测性，$\bar{P}_L$ 的定义域的积 σ-代数和 H 上$\mathcal{B}$的博雷尔 σ-代数必须可以在某种方式下被扩张．首先，$\mathcal{B}$能扩大．存在普遍可测集的 σ-代数$\mathcal{U}$，它对$\mathcal{B}$上所有法则的完备化都是可测的．(它比$\mathcal{B}$严格大，例如，就像 13.2 节中所述的它包含所有解析集.)尽管 H 上没有给出概率测度，但是如果选择一个定义在$\mathcal{B}$上的概率测度(如$\mathcal{U}$)和其他集合使其是完备的，那么它的可测集的 σ-代数$\mathcal{M}(\mu)$将会包含$\mathcal{U}$．同样，在$\overline{\mathbb{R}}^H\times H$ 上将得到一个积测度 $\bar{P}_L\times\mu$，对它我们可以取一个完备化．从而如果$(f,\ t)\mapsto f(t)$是可测的，那么对$\mathcal{M}(\mu)\bar{P}_L$-几乎所有的 f 是可测的(顺便说一下，如果积测度的一般完备化用 Bledsoe 和 Morse(1955)的方法扩张，这仍然是成立，这种方法是在 4.4 节的习题 14 中讨论的)．对 H 上的任意 σ-代数$\mathcal{S}$，令$\mathcal{L}^0(H,\ \overline{\mathbb{R}},\ \mu)$为从 H 映射到$\overline{\mathbb{R}}$ 的所有函数的集族，对$\mathcal{S}$是可测的．令$\mathcal{S}^0(H,\ \overline{\mathbb{R}},\ \mu)=\mathcal{S}^0(H,\ \overline{\mathbb{R}},\ \mathcal{M}(\mu))$．

回想由 $\mu^*(A):=\inf\{\mu(B):\ B\supset A,\ B\text{ 可测}\}$定义的外测度，和由 $\mu_*(A):=\sup\{\mu(B):\ B\subset A,\ B$ 可测$\}$定义的内测度．现在可以给出本附录的一个主要事实．

E.2 命题 在紧空间$\overline{\mathbb{R}}^H$ 上，令 P_L 是类正态过程的法则，$\bar{P}_L$ 是其完备正则化扩张，那么在连续统假

设下，H 上博雷尔可测或普遍可测函数空间对 $\bar{P}_L$ 内测度为 0：

$$(\bar{P}_L)_*(\mathcal{L}^0(H,\overline{\mathbb{R}},\mathcal{B})) = (\bar{P}_L)_*(\mathcal{L}^0(H,\overline{\mathbb{R}},\mathcal{U})) = 0.$$

事实上，在 H 上存在一个完备法则 μ，使得

$$(\bar{P}_L)_*(\mathcal{L}^0(H,\ \overline{\mathbb{R}},\ \mu)) = 0.$$

因此，$(f,\ t)\mapsto f(t)$ 对于积测度 $\bar{P}_L\times\mu$ 的完备化是不可测的.

证明 令 μ 是 H 上的测度，H 是 $\sum_n y_n e_n$ 的分布，其中 y_n 是满足 $\mathcal{L}(y_n)=N(0,\ n^{-3/2})$ 的独立随机变量.（因为变量的和收敛，所以这就给出了 H 上的一个法则.）令 $\mathcal{V}$ 是定义在 H 上的所有函数 V 的集合，函数 V 满足对每个 $f\in H$，$V_f:=V(f)$ 是一个在 $\mathbb{R}$ 上有有理端点的非空开区间. 对任意可测集 $A\subset\Omega$，有限集 $F\subset H$ 和 $V\in\mathcal{V}$，令

$$A_{F,V} := \{z\in A: L(f)(z)\in V_f, \forall f\in F\}.$$

命题 E.2 的证明主要基于以下事实.

E.3 引理 在连续统假设下，对每一个满足 $P(A)>0$ 的可测集 $A\in\Omega$，存在一个满足 $\mu^*(S)=1$ 的集合 $S\subset H$，使得对每一个有限集 $F\subset S$ 和 $V\in\mathcal{V}$，有

$$P(A_{F,V}) > 0. \qquad (*)$$

证明 注意到为使结论成立，F 和 S 必须是线性独立的（对于有限的线性组合）. 由连续统假设，取一个良序集 $(J,\ <)$，其中 J 的势为 c（$\mathbb{R}$ 的势），且使得对每个 $\alpha\in J$，$\{\beta\in J:\ \beta<\alpha\}$ 是可数的. 由于 H 是可分的，对其开集，它存在一个可数基（命题 2.1.4）. 因此可数基的所有子集组成的集族的势也为 c. 那么 H 中恰好存在 c 个开集和 c 个闭集. 令满足 $\mu(Y)>0$ 的所有闭集 Y 组成的集族的指标为 $\{Y_\alpha:\ \alpha\in J\}$. 将递归地定义集合 $S:=\{s_\alpha:\ \alpha\in J\}$.

对 $n=1,\ 2,\ \cdots$，令 $\mathcal{C}_n$ 是所有笛卡儿积 $\prod_{1\le k\le n}[a_k,a_k+1/2^n]\times\prod_{k>n}\mathbb{R}\subset\Omega$ 组成的集族，其中每个 a_k 可能都有 $-n$，$-n+1/2^n$，$-n+2/2^n$，$\cdots$，$n-1/2^n$ 中的任意值. 令 $\mathcal{F}_n:=\mathcal{C}_n\cup\{D_n\}$，其中 D_n 是 $\mathcal{C}$ 的并的补集. 令 $\mathcal{A}_n$ 是由 $\mathcal{F}_n$（或 $\mathcal{C}_n$）生成的代数. 那么 $\mathcal{A}_n$ 是一个有限代数的递增序列，并且它的并生成了 Ω 的博雷尔 σ-代数. 对每个集合 $A\in\mathcal{F}_n$，我们都有 $P(A)>0$，并且对任意的博雷尔集合 $B\subset\Omega$，在每个 $A\in\mathcal{F}_n$ 上，$P(B\mid\mathcal{A}_n)=P(B\cap A)/P(A)$. 根据鞅收敛定理（10.5.4），$P(B\mid\mathcal{A}_n)\to1$ 在 B 的几乎所有点上都成立. 使上式成立的 B 的点称为 B 的稠密点（对 $\mathcal{A}_n$）. 对任意的 $z\in\Omega$，令 $E^n(z)$ 是 $\mathcal{F}_n$ 中的唯一集合，$z\in E^n(z)$. 那么 z 是 B 的稠密点当且仅当 $z\in B$ 并且

$$\lim_{n\to\infty}P(E^n(z)\cap B)/P(E^n(z)) = 1.$$

一个实数序列 $\{r_n\}$ 称为递归的（recurrent），当且仅当对每一个非空开区间 $U\subset\mathbb{R}$，对无穷多个 n，有 $r_n\in U$. 对任意有限的 N，显然，这个性质不依赖于 r_1，$\cdots$，r_N. 令 $A(k)$ 是整数 $n=2^k$，$\cdots$，$2^{k+1}-1$（$k=0,\ 1,\ \cdots$）的集合. 对 $n=0,\ 1,\ \cdots$，令 $k(n)$ 是唯一的 $k=0,\ 1,\ \cdots$，使得 $n\in A(k)$. 为了继续引理 E.3 的证明，还需要下面两个引理.

E.4 引理 对 $n=0,\ 1,\ \cdots$，令 $y_n:=(-1)^{k(n)}\omega_n/n^{3/4}$，那么对 P 几乎必然地有部分和 $S_n:=S(n):=\sum_{1\le j\le n}y_j\omega_j$ 形成一个递归序列.

证明 令 $Z_k:=(-1)^k\sum_{j\in A(k)}\omega_j^2/j^{3/4}$，那么 $EZ_k=(-1)^k\sum_{j\in A(k)}j^{-3/4}$，当 $k\to\infty$ 时，通过与 $x^{-3/4}$ 从 2^k 到 2^{k+1} 和从 2^k-1 到 $2^{k+1}-1$ 的积分的比较知，它渐近收敛于 $(-1)^k2^{(k+8)/4}(2^{1/4}-1)$. 因为 $E\omega_j^4=3$（由

命题 9.4.2c)，所以 Z_k 的方差为 $2\sum_{j\in A(k)} j^{-3/2}$，当 $k\to\infty$ 时，它渐近收敛于一个 $2^{-k/2}$ 的常数倍．所以由切比雪夫不等式和博雷尔–坎泰利引理知，对所有足够大的 k，几乎必然有 $|Z_k - EZ_k| < 1$．那么 Z_{k+1}/Z_k 收敛于 $-2^{-1/4} < -1$，$Z_{2k+1} + Z_{2k} \to -\infty$，而 $Z_{2k} + Z_{2k-1} \to +\infty$．当 $k\to\infty$ 时，$S(2^{2k}-1)\to -\infty$，而 $S(2^{2k+1}-1)\to +\infty$．另一方面，当 $j\to\infty$ 时，$y_j w_j \to 0$ a. s.，换句话说，再次从 $E\omega_j^4 = 3$、切比雪夫不等式和博雷尔–坎泰利引理可得，$\omega_j^2/j^{3/4}\to 0$．所以当 $k\to\infty$ 时，逐步小地从 $S(4^k)$ 到 $S(2^{2k+1})$，S_n 在 $\mathbb{R}$ 上变得越来越密集，引理 E.4 得证． □

$S(n)$ 是递归的所有 $\omega := \{w_j\}_{j\geq 1}$ 的集合 $\mathcal{R}$ 是可测的（取具有有理端点的区间），且 $P(\mathcal{R}) = 1$．令 $\mathcal{M}$ 为满足 $\sum_j \omega_j{}^2/j^{3/2} < \infty$ 的所有 ω 的集合，并且使得对所有足够大的 N，有 $\max_{j\leq N}|\omega_j| \leq N$．因为由引理 12.1.6 知，$\sum_N NP(|\omega_1| > N) < \infty$，故 $P(\mathcal{M}) = 1$．对每个有限集 $F\subset H$ 和 $V\in\mathcal{V}$，令

$$T_{F,V} := \{y \in H: A_{F,V}\cap\mathcal{R}\cap\mathcal{M}\text{的某些稠密点},\omega,$$
$$y_n = (-1)^{k(n)}\omega_n n^{-3/4},\text{对所有的 } n \geq m \text{ 和某些 } m = m(y)\}.$$

E.5 引理 *$T_{F,V}$ 是可测的，并且如果 $P(A_{F,V}) > 0$，那么 $\mu(T_{F,V}) = 1$.*

证明 几乎处处定义在 Ω 到 H 上的 1－1 可测函数 $\omega \mapsto \sum_n (-1)^{k(n)}\omega_n n^{-3/4} e_n$ 是保测的，把 P 代入 μ．$A_{F,V}\cap\mathcal{R}\cap\mathcal{M}$ 的所有稠密点的集合对于 P 几乎是全部 $A_{F,V}$ 的集合，所以它的概率为 $P(A_{F,V})$．如果在 $T_{F,V}$ 的定义中固定 $m = 1$，就会得到一个概率为 $P(A_{F,V})$ 的可测集．假设我们以具有相同概率包含在它内的博雷尔集合替换 $A_{F,V}\cap\mathcal{R}\cap\mathcal{M}$ 的稠密点集，那么对 $T_{F,V}$ 的定义中每一个固定的 m，由定理 13.2.1，我们得到一个解析集 $C_{F,V,m}$，并且由定理 13.2.6 知，它是普遍可测的．这些集合随 m 递增．因此对每一个 n，它们的并 $C_{F,V}$ 独立于 $y_1,\cdots,y_n$．另一方面，$C_{F,V}$ 是独立于 μ 的变量序列 $\{y_j\}_{j\geq 1}$ 的一个函数．我们有 $\mu(C_{F,V}) > 0$，所以由科尔莫戈罗夫 0－1 律(8.4.4)知，$\mu(C_{F,V}) = 1$．由于 $C_{F,V}\subset T_{F,V}$，$\mu(T_{F,V}) = 1$ 其中 $T_{F,V}$ 是可测的，由假设知，μ 是完备的，引理 E.5 得证． □

现在继续引理 E.3 的证明，为了具体化 S，给定 $S_\alpha := \{s_\beta:\ \beta < \alpha\}$，使得对 S_α 的任意有限子集 F 都有 $P(A_{F,V}) > 0$ 且 $V\in\mathcal{V}$，令 $T_\alpha := \bigcap_{F,V} T_{F,V}$，其中 F 取遍所有 S_α 的有限子集，V 取遍 $\mathcal{V}$．那么由引理 E.5 知，$\mu(T_\alpha) = 1$，因为 S_α 是可数的，F 的可能集合和对 $f\in F$ 的 V_f 都是可数的．因此，可以选取 $s_\alpha\in T_\alpha\cap Y_\alpha$．所以 S 的递归定义是完备的．由于 S 与每一个 Y_α 是相交的，并且博雷尔集合上的任意测度都是闭正则的(定理 7.1.3)，故 $\mu^*(S) = 1$．

现在，引理 E.3 中的(*)式每次仅依赖于 S 的有限多个元素．对某个 $\alpha\in J$，S 的任意有限子集均包含在 S_α 中．所以只需证明对每一个 α，S_β 满足(*)式．这将递归地被证明，假设对所有的 $\beta < \alpha$，S_β 满足(*)式．如果没有最大的 $\gamma < \alpha$，那么对某个 $\beta < \alpha$，S_α 的任意有限子集包含在 S_β 中，根据归纳假设知，这些结论对以上都成立．所以可以假设存在最大的 $\beta < \alpha$，那么 $S_\alpha = S_\beta\cup\{s_\beta\}$．

令 F 是 S_α 的一个有限子集，那么如果 $F\subset S_\beta$，则(*)式成立，所以为了证明(*)，对 S_β 的一个有限子集 G 和 $t = s_\beta$，可以假设 $F = G\cup\{t\}$．对任意的 $V\in\mathcal{V}$，需要证明 $P(A_{F,V}) > 0$，给定 $P(A_{G,V}) > 0$．

因为 $t = s_\beta\in T_{G,V}$，所以存在 $A_{G,V}\cap\mathcal{R}\cap\mathcal{M}$ 的一个稠密点 z，使得对某个 K，

$$t_j = (-1)^{k(j)} z_j j^{-3/4},\quad \forall j \geq K,\text{且对所有足够大的 } N,$$
$$|z_j| \leq N,\quad j = 1,\cdots,N. \tag{‡}$$

令 $V_t=(a,\ b)$，$\varepsilon:=(b-a)/3>0$. 存在一个 $M>K$，使得 $P\{|\sum_{j\geq M}t_j\omega_j|\geq\varepsilon/2\}<1/2$，因为 $\sum_{j\geq M}t_j\omega_j$ 是均值为零的正态随机变量，且当 $M\to\infty$ 时，方差 $\sum_{j\geq M}t_j^2\to 0$.

回想在引理 E.3 证明的第二段中所定义的 $E^n(z)$. 由独立性，对所有的 $N\geq M$，

$$P(E^N(z)\cap\{\left|\sum_{j>N}t_j\omega_j\right|<\varepsilon/2\})>P(E^N(z))/2.$$

另一方面，因为 z 是 $A_{G,V}$ 的一个稠密点，对所有足够大的 N，

$$P(A_{G,V}\cap E^N(z))>P(E^N(z))/2.$$

那么对

$$D_N:=A_{G,V}\cap E^N(z)\cap\{\left|\sum_{j>N}t_j\omega_j\right|<\varepsilon/2\},P(D_N)>0 \qquad (\ddagger)$$

根据 Cauchy-Bunyakovsky-Schwarz 不等式，对 D_N 中任意的 $\omega:=\{\omega_j\}_{j\geq 1}$，

$$\left|L(t)-\sum_{1\leq j\leq N}t_jz_j\right|\leq S_1+S_2$$

其中

$$S_1:=\left|\sum_{j>N}t_j\omega_j\right|<\varepsilon/2,$$

$$S_2:=\left|\sum_{1\leq j\leq N}t_j(z_j-\omega_j)\right|\leq\|t\|\left(\sum_{1\leq j\leq N}(z_j-\omega_j)^2\right)^{1/2}.$$

现在 $\|t\|$ 固定，$\omega\in E^N(z)$，自($\ddagger$)式满足 $|z_j|\leq N$ 推得，对 $j=1,\ \cdots,\ N$，$|\omega_j-z_j|<1/2^N$，所以对 $N\geq N_0\geq M$，$S_2\leq\|t\|N^{1/2}/2^N<\varepsilon/2$，其中 N_0 不依赖于 ω. 对这样的 N 和任意的 $\omega\in D_N$，

$$\left|L(t)(\omega)-\sum_{1\leq j\leq N}t_jz_j\right|<\varepsilon.$$

现在序列 $\sum_{1\leq j\leq N}t_jz_j$ 是递归的，因为 t 和 z 是通过式($\ddagger$)相联系的，且 $z\in\mathcal{R}$. 所以存在一个 $N\geq N_0$，使得 $\sum_{1\leq j\leq N}t_jz_j$ 处于 V_t 的三等分位置，即区间 $(a+\varepsilon,\ a+2\varepsilon)$ 中. 对这样的 N，我们得到

$$P(A_{F,V})\geq P(D_N\cap\{L(t)\in(a,b)\})=P(D_N)>0,$$

引理 E.3 得证. □

继续命题 E.2 的证明：令 C 是 $\overline{\mathbb{R}}^H$ 的一个紧子集，$\bar{P}_L(C)>0$，且 $C\subset\mathcal{L}^0(H,\ \overline{\mathbb{R}},\ \mu)$，那么对满足 $\bar{P}_L(C)=\bar{P}_L(C_1)=P_L(C_1)$ 的紧贝尔集 C_1，有 $C\subset C_1$，其中由正则性知，$\bar{P}_L(C)=\operatorname{int}\{\bar{P}_L(U)$：$U$ 为开集，$U\supset C\}$. 对于 $n=1,\ 2\cdots$，取一个满足 $C\subset U_n$ 和 $\bar{P}_L(U_n\setminus C)<1/n$ 的开集 U_n. 由定理 2.6.2 和 Urysohn 引理(2.6.3)知，存在一个满足 $0\leq f_n\leq 1$ 的连续函数 f_n，在 C 上 $f_n=0$，在 U_n 外 $f_n=1$. 那么 $D_n:=f_n^{-1}[0,\ 1/2]$ 是紧贝尔集，正如期望的它们的交是紧贝尔集 C_1.

正如在定理 7.7.1 后面的例子中所表述的，对可数集 $Y\subset H$ 中的 y，$f\in C_1$ 是否仅依赖于 $f(y)$. 令 η 是从 Ω 映射到 $\mathbb{R}^Y$ 由 $\eta(\omega)(y):=L(y)(\omega)$ 定义的函数. 对每个 y，这个函数对 P-几乎所有的 ω 是有定义的，并且由于 Y 是可数的，对 P-几乎所有的 ω，它被定义到 $\mathbb{R}^Y$ 上. 定义后，η 是 P 可测的. 现在 $C_1=\pi^{-1}(C_2)$，其中 C_2 是 $\overline{\mathbb{R}}^Y$ 上的一个博雷尔集合，π 是从 $\overline{\mathbb{R}}^H$ 映上到 $\overline{\mathbb{R}}^Y$ 的一个自然投影. 从定义得知，$P(\eta^{-1}(C_2))=P_L(C_1)$. 把引理 E.3 应用到 $A=\eta^{-1}(C_2)$ 可得一个满足 $\mu^*(S)=1$ 的集合 $S\subset H$，使得对任意有限的 $F\subset S$ 和 $V\in\mathcal{V}$，都有 $P(\eta^{-1}(C_2)_{F,V})>0$. 等价地，

$$P_L\{\varphi\in C_1:\varphi(f)\in V_f,\forall f\in F\}>0.$$

由 C_1 的选取，则有

$$\bar{P}_L\{g \in C: g(f) \in V_f, \forall f \in F\} > 0.$$

因为 C 对于逐点收敛来说是紧的，所以可得从 S 映射到 $\overline{\mathbb{R}}$ 的所有函数都是 C 中函数的限制.

对任意的 μ-可测集 $E \subset H$，令 $v(E \cap S) := \mu^*(E \cap S)$. 那么由定理 3.3.6 知，$v$ 是一个可数可加的概率测度. 对每一点 $p \in S$，$v(\{p\}) \leqslant \mu(\{p\}) = 0$. 所以由连续统假设和定理 C.1 知，$v$ 不是定义在 S 的所有子集上. 所以并不是每一个从 S 映射到 $\overline{\mathbb{R}}$ 的函数都能扩张成一个从 H 映射到 $\overline{\mathbb{R}}$ 的 μ-可测函数. 这与 $C \subset \mathcal{L}^0(H, \overline{\mathbb{R}}, \mu)$ 矛盾. 所以 $\mathcal{L}^0(H, \overline{\mathbb{R}}, \mu)$ 的内测度 $\bar{P}_L$ 是 0. □

现在将会注意到正则性扩张对合适的过程可以做些什么. 令 $E[0, 1]$ 是 $[0, 1]$ 上所有实值函数的集合，使得对于每个 $t \in [0, 1]$，左极限 $f(t^-) = \lim_{u \uparrow t} f(u)$ (除了 $t=0$ 外) 和右极限 $f(t^+) = \lim_{u \downarrow t} f(u)$ (除了 $t=1$ 外) 都存在且有限. 一个曾经被深入研究过的更小的空间是满足 $f(t) = f(t^+)$ $(0 \leqslant t < 1)$ 的函数 $f \in E[0, 1]$ 的子空间 $D[0, 1]$. $D[0, 1]$ 中的函数都是右连续的且具有左极限，就像 $\mathbb{R}$ 上概率法则 P 的累积分布函数 F，$F(x) := P((-\infty, x])$. 如果一个过程在 $E[0, 1]$ 上有样本函数，那么如命题 E.2 中的病态性就不会发生了.

E.6 定理 (E. Nelson) 令 $I := [0, 1]$，那么 $E[0, 1]$ 是 $\overline{\mathbb{R}}^I$ 中的博雷尔集，事实上，它是一个 $K_{\sigma\delta}$——$\mathbb{R}^I \subset \overline{\mathbb{R}}^I$ 中紧集的可数并的一个可数交. $E[0, 1]$ 上每一个函数在 I 上都是博雷尔可测的. 令 x 是 $T = I$ 上的一个随机过程，且概率空间 Ω 使得对几乎所有的 $\omega \in \Omega$，函数 $t \mapsto x_t(\omega)$ 在 $E[0, 1]$ 中. 那么对 $\overline{\mathbb{R}}^T$ 上 x 的法则的正则性扩张 $\bar{P}_x$，$\bar{P}_x(E[0, 1]) = 1$. 因此，博雷尔可测函数的集合、普遍可测函数，或者可测函数对 $[0, 1]$ 上的一个固定测度，每一个都有内测度 1.

证明 如果 u 或者 v 是 $\pm\infty$，若 $u = v$，令 $|u - v| := +\infty$，否则 $|u - v| := -\infty$. 为了证明 $E[0, 1]$ 是一个 $K_{\sigma\delta}$，对 $k, n = 1, 2 \cdots$ 令

$$U_{nk} := \{f \in \overline{\mathbb{R}}^I : \exists x_j, 0 \leqslant x_1 < x_2 < \cdots < x_n \leqslant 1, |f(x_j) - f(x_{j-1})| > \frac{1}{k}, j = 2, \cdots, n\},$$

$V_n := \{f \in \overline{\mathbb{R}}^I : \exists x, 0 \leqslant x \leqslant 1, |f(x)| > n\}$，这两个都是开集. 令 $W_{nk} := U_{nk} \cup V_n$. $E[0, 1]$ 上的函数一定是有界的. 因为对每一个 V_n^c，有一个 $K_{\sigma\delta}$，从而有 $E[0, 1] = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} W_{nk}^c$，从而 W_{nk}^c 是紧的. 对每一个 k，$W_{1k} \supset W_{2k} \supset \cdots$ 所以 $W_{1k}^c \subset W_{2c}^c \subset \cdots$.

如果 $f \in E[0, 1]$，那么除了至多在一个可数集上外，f 是连续的，从而是博雷尔可测的.

如果对某个 k 和 $\delta > 0$，$\bar{P}_x(W_{nk}^c) < 1 - \delta$ 对所有的 n 成立，那么存在满足 $P_x(B_n) = \bar{P}_x(W_{nk}^c)$ 的贝尔集 $B_n \supset W_{nk}^c$. 令 $B := \bigcup_n B_n$，一个包括满足 $P_x(B) \leqslant 1 - \delta$ 的所有 W_{nk}^c 的贝尔集. 那么 B 仅依赖于可数集 T 上的坐标，并且对所有的 n，$B^c \subset W_{nk}$，这与 x_t 在 $E[0, 1]$ 上存在样本函数矛盾. 剩下的证明参见 Nelson(1959, 定理 3.4). □

注：正如刚刚陈述的，根据定理 E.6 所有内测度为 1 的可测函数集合在希尔伯特空间上对类正态过程 L 具有内测度 0，正如命题 E.2 所证明的. 定义在 $[0, 1]$ 上的很多随机过程能在 $D[0, 1]$ 上从而在 $E[0, 1]$ 上具有样本函数 $t \mapsto x_t(\omega)$. 一个例子就是经验分布函数

$$F_n(t) := \sum_{1 \leqslant j \leqslant n} 1_{[0, t]}(X_j)$$

其中 X_j 是取值于 $[0, 1]$ 的随机变量，或者特殊地，是具有分布函数 F 的独立同分布随机变量. 那么正规化函数 $n^{1/2}(F_n - F)$ 也在 $E[0, 1]$ 内(如果 F 是均匀分布函数，对于 $0 \leqslant t \leqslant 1$，$F(t) = t$，那么

当 $n \to \infty$ 时，这些函数依 L 收敛到布朗桥，正如在 12.1 节中所提到的）．另一个例子是满足一些常用正则性条件的马尔可夫过程．参见 Blumenthal 和 Getoor(1968，p. 45—46)．

另一方面，$D[0, 1]$ 是 $\overline{\mathbb{R}}^I$ 上的一个高度不可测子集，正如下面所证明的．令 $M_- := \{1_{[x,1]}: 0 \leqslant x \leqslant 1\}$，$M_+ := \{1_{(x,1]}: 0 \leqslant x < 1\}$，那么 $M := M_- \cup M_+$ 在 $I^I \subset \overline{\mathbb{R}}^I$ 上是紧的，并且 $D[0, 1] \cap M = M_-$．但是对 I^I 上的任意非原子正则博雷尔测度 γ，M_- 和 M_+ 的内测度都为 0，因为它们当中的任一个紧子集都是可数的．因此如果 $\gamma(M) > 0$，那么 M_- 和 M_+ 对 γ 是不可测的．（注意：M_- 是可能函数 $F_1(\cdot)$ 的集合．）所以，从 $[0, 1]$ 映射到 I^I 的映射 $x \mapsto 1_{[x,1]}$ 的值域在 I^I 上不是博雷尔可测的或者普遍可测的，尽管具有博雷尔 σ-代数的映射在 I^I 上是可测的，因为 I^I 中任何开集的逆像都是区间的并，这些区间或者是非退化的，或者是 $\{0\}$．这种并是博雷尔集，因为它是开集和可数集的并．回想相比之下，从 I 映射到具有博雷尔 σ-代数的 I^I 上的 $x \mapsto 1_{\{x\}}$ 是不可测的，或者仅对定义域 I 的所有子集的 σ-代数是可测的，就像任意函数那样（命题 4.2.3）．它们的值域是博雷尔集，对于紧集 K 具有形式 $K \setminus \{0\}$．

注释

命题 E.2 及其证明都来自于 Dudley(1972，1973)；Vaclav Fabian 就有关修正(1973)给出了非常有用的建议．命题 E.2 解决了 Kakutani（1943）和 Doob(1947)提出的一个问题．Bledsoe 和 Morse(1955)定义了它们的扩充的积测度，扩充的积测度对每一个集合都给出了测度 0，这些集合的示性函数以每一顺序的累次积分都为 0．Nelson(1959)证明了像定理 E.6 一样更一般的正结果，表明各类函数都是 $K_{\sigma\delta}$ 的．例如，对 T 上的某一测度，对几乎所有的 $x_t(\omega)$，如果 $t \mapsto x_t(\omega)$ 在 t 处是连续的，那么它是充分的．因此命题 E.2 表明，这样的结果不能进一步扩张到像类正态这样的过程上，这里 $t \mapsto x_t$ 是依概率连续的，但是对固定的 ω 是不连续的．Tjur(1980，10.9.4)在最后一段给出了 M_- 的非可测性的说明．同样参见 Dudley(1990)．

参考文献

Bledsoe, Woodrow W., and Anthony P. Morse (1955). Product measures. *Trans. Amer. Math. Soc. 79*: 173–215.

Blumenthal, R. M., and R. K. Getoor (1968). *Markov Processes and Potential Theory*. Academic Press, New York.

Bourbaki, N. (1952–1969). *Intégration*, Chaps. 1–4 (1952), 2d ed. (1965); Chap. 5 (1956), 2d ed. (1965); Chap. 6 (1959); Chaps. 7–8 (1963); Chap. 9 (1969). Hermann, Paris.

Doob, Joseph L. (1947). Probability in function space. *Bull. Amer. Math. Soc. 53*: 15–30.

Dudley, R. M. (1972, 1973). A counterexample on measurable processes. *Proc. 6th Berkeley Symp. Math. Statist. Prob.* (Univ. of Calif. Press) 2: 57–66; Correction, *Ann. Probability 1*: 191–192.

——— (1990). Nonmetric compact spaces and nonmeasurable processes. *Proc. Amer. Math. Soc. 108*: 1001–1005.

Halmos, Paul R. (1950). *Measure Theory*. Van Nostrand, Princeton.

Kakutani, Shizuo (1943). Notes on infinite product spaces, II. *Proc. Imp. Acad. Tokyo 19*: 184–188.

Nelson, Edward (1959). Regular probability measures on function space. *Ann. Math. 69*: 630–643.

Tjur, Tue (1980). *Probability Based on Radon Measures*. Wiley, New York.

名词索引

索引中的页码为英文原书页码，与书中页边标注的页码一致．

A

B

C

D

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z

符 号 索 引

用希腊和拉丁字母表示的符号：

不是字母的符号，定义如下：



华章数学译丛

数学文化
数学简史（Katz，英）

微积分
高等微积分（Fitzpatrick，中、英）
微积分及其应用（Bittinger，中）
大学微积分（Hass，中）

数学分析
数学分析原理（Rudin，中、英）
数学分析（Apostol，中、英）
纯数学教程（Hardy，英）
泛函分析（Rudin，中、英）
实分析与复分析（Rudin，中、英）
实分析（Royden，中、英）
实分析和概率论（Dudley，中、英）
复分析（Ahlfors，中、英）
复变函数及应用（Brown，中、英）
复分析基础及工程应用（Saff，中、英）
三角级数（Zygmund，英）

调和分析
调和分析导论（Katznelson，英）
逼近论教程（Cheney，英）
小波基础及应用教程（Mix，中）
小波与小波变换导论（Burrus，中、英）
小波分析及其应用（孙延奎，编写）
傅里叶分析与小波分析导论（Pinsky，英）
时频变换与小波变换导论（钱世锷，英）

代数
线性代数（Jain，英）
线性代数（Leon，中、英）
线性代数及其应用（Lay，中）
代数（Isaacs，英）
代数（Artin，中、英）
抽象代数基础教程（Rotman，中、英）
高等近世代数（Rotman，中）
矩阵分析（Horn，中）
同调代数导论（Weibel，英）

几何、拓扑
曲线与曲面的微分几何（do Carmo，中、英）
微分几何及其应用（Oprea，中、英）
分形分析（Kigami，英）
拓扑学（Munkres，中、英）

数学建模
数学建模方法与分析（Meerschaert，中）
数学建模（Giordano，中、英）

微分方程
实用偏微分方程（Haberman，中、英）
偏微分方程教程（Asmar，中、英）
微分方程与边界值问题（Zill，中、英）
动力系统导论（Robinson，中、英）
流体动力学导论（Batchelor，英）

计算数学
数值方法和MATLAB实现与应用（Reckten-wald，中）
数值分析（Kincaid，中、英）
数值方法（金一庆，编写）
计算机数值计算方法及程序设计（周煦，编写）
科学计算导论：使用MATLAB的矩阵向量方法（Van Loan，英）
MATLAB数值计算（Moler，中）
具体数学：计算机科学基础（Graham，中、英）

概率统计
概率论与数理统计（陈方樱，编写）
概率论基础教程（Ross，中）
概率统计（Stone，英）
概率论及其在投资、保险、工程中的应用（Bean，英）

华章数学译丛

概率与计算（Mitzenmacher，中）
贝叶斯方法（Leonard，英）
抽样理论与方法（Govindarajulu，英）
数理统计与数据分析（Rice，英）
应用回归分析和其他多元方法（Kleinbaum，英）
多元数据分析（Lattin，英）
预测与时间序列（Bowerman，英）
时间序列分析的小波方法（Percival，中、英）
随机过程导论（Kao，英）
试验者的统计学（Box，中）
理工科概率统计（Walpole，中）
统计学（Mendenhall，中）

离散数学

离散数学（陈国勋，编写）
离散数学导学（Simpson，中）
离散数学及其应用（Rosen，中、英）
离散数学（Dossey，中、英）
离散数学及其应用（徐凤生，编写）

组合数学

组合数学教程（van Lint，中、英）
组合数学（Brualdi，中、英）
应用组合数学（Roberts，中、英）
计数组合学：卷1、卷2（Stanley，英）
图论导引（West，中、英）
图论（Tutte，英）
网络流：理论、算法与应用（Ahuja，英）

数论

初等数论及其应用（Rosen，中、英）
数论概论（Silverman，中、英）

数理逻辑

应用逻辑（Nerode，中、英）

金融数学

金融数学（Stampfli，中、英）
数理金融初步（Ross，中、英）
金融时间序列分析（Tsay，中）

运筹学

数学规划导论（Walker，英）
线性规划导论（Vaserstein，中、英）

数学软件

LATEX实用教程（Kopka，英）
MATLAB 7及工程问题解决方案（Etter，中）
SAS统计分析及应用（黄燕，编写）